

# **Kopplungen und nicht intendierte Effekte in Bildungssystemen**

Empirische Untersuchungen aus Kanada und Deutschland sowie  
eine steuerungstheoretische Problematisierung

Abhandlung (kumulative Dissertation)  
zur Erlangung der Doktorwürde  
der Philosophischen Fakultät  
der  
Universität Zürich

Vorgelegt von  
Daniela Julia Jäger

Angenommen im Herbstsemester 2013  
auf Antrag der Promotionskommission:  
Prof. Dr. Katharina Maag Merki  
(hauptverantwortliche Betreuungsperson)  
Prof. Dr. Lucien Criblez

Zürich, 2013

## Inhalt

- Synopse: Daniela J. Jäger: Kopplungen und nicht intendierte Effekte in Bildungssystemen. Empirische Untersuchungen aus Kanada und Deutschland sowie eine steuerungstheoretische Problematisierung.
- Daniela J. Jäger (2014): Zwischen Empowerment und Kontrolle. Die praktische Umsetzung des New Public Management und Professioneller Lerngemeinschaften in Kanada. Eine Fallstudie. *Zeitschrift für Pädagogik*, 60 (2), 303-324.
- Daniela J. Jäger, Katharina Maag Merki, Britta Oerke & Monika Holmeier (2012): Statewide low-stakes tests and a teaching to the test effect? An analysis of teacher survey data from two German states. *Assessment in Education: Principles, Policy & Practice*, 19 (4), 451-467.
- Daniela J. Jäger, 2012: Herausforderung Zentralabitur: Unterrichtsinhalte variieren und an Prüfungsthemen anpassen. In: Maag Merki, Katharina (Hrsg.): *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland*. Wiesbaden: Springer VS, S. 179-205.
- Daniela J. Jäger, 2012: Schulklima, Selbstwirksamkeit und Arbeitszufriedenheit aus Sicht der Lehrpersonen und Schüler/-innen in Hessen und Bremen. In: Maag Merki, Katharina (Hrsg.): *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland*. Wiesbaden: Springer VS, S. 65-93.
- Anhang (Erklärung, Lebenslauf)

Synopse: Daniela J. Jäger: Kopplungen und nicht intendierte Effekte in Bildungssystemen. Empirische Untersuchungen aus Kanada und Deutschland sowie eine steuerungstheoretische Problematisierung.

## *Inhaltsverzeichnis*

1	Einleitung .....	1
2	Governance, Steuerung, Kopplungen, Kopplungsmechanismen und (nicht) intendierte Effekte .....	6
2.1	Governance und Steuerung .....	6
2.2	Kopplungen und Kopplungsmechanismen .....	8
2.3	Intendierte und nicht intendierte Effekte .....	12
3	Kopplungsmechanismen und nicht intendierte Effekte .....	13
3.1	Ursachen nicht intendierter Effekte .....	13
3.2	Reduzierung nicht intendierter und Forcierung intendierter Effekte .....	15
3.2.1	Kopplungsmechanismen als Antwort auf fehlende Information und „hidden information“? .....	15
3.2.2	Kopplungsmechanismen als Antwort auf „hidden action“? .....	16
4	Was ist ein Bildungssystem? .....	18
4.1	Die Schule als Organisation .....	18
4.2	Funktionen von Schule und Bildungssystem .....	20
4.3	Spezifika des Bildungssystems .....	21
4.4	Implikationen für die Governance – Empirische Befunde .....	23
5	Zwischenfazit und übergeordnete Fragestellungen zu den Studien .....	24
6	Erwartete Kopplungsmechanismen in der best-case Studie .....	25
6.1	Fragestellungen der Studie A .....	25
6.2	Methoden der Studie A .....	26
6.3	Ergebnisse der Studie A .....	27
7	Erwartete, nicht intendierte Effekte in den Studien zum Zentralabitur .....	29
7.1	Fragestellungen der Studien B, C und D .....	30
7.2	Methoden der Studien B, C und D .....	31
7.3	Ergebnisse der Studien B, C und D .....	35
8	Diskussion .....	37
8.1	Kopplungsmechanismen in einem best-case .....	38
8.2	Nicht intendierte Effekte des Zentralabiturs .....	41
9	Fazit und Ausblick .....	43

## Literaturverzeichnis

### *Tabellenverzeichnis*

Tabelle 1: Kopplungsmechanismen.....	11
Tabelle 2: Eingesetzte Methoden der Studie A pro Untersuchungsbereich .....	26
Tabelle 3: Charakteristika von NPM und PLG.....	28
Tabelle 4: Erfolgsversprechende Faktoren? .....	29
Tabelle 5: Fragestellungen der Studien B, C und D im Überblick .....	32
Tabelle 6: Stichprobe und Rücklauf der Lehrpersonen.....	32
Tabelle 7: Stichprobe und Rücklauf der SchülerInnen .....	33
Tabelle 8: Methodische Differenzen zwischen den Studien B, C und D .....	35

## **Dank**

Es ist es geschafft - und hier ist endlich die Chance episch-pathetisch und voll und ganz subjektiv zu danken. In diesem Sinne:

Herzlichen Dank an Jan für das gemeinsame Durchdenken theoretischer Zusammenhänge und das "zurück-auf-die-Spur-setzen" bei den Wanderungen durch begriffliches Gestrüpp. Außerdem für die schöne gemeinsame Ausgleichszeit! Letzteres gilt auch unserem Sohn und meinen Freundinnen und Freunden.

Danke - meiner Mutter und meinem Vater für die Lust zu diskutieren und den Mut neue Dinge auszuprobieren – ohne diese Zugaben hätte ich keine Freude an der wissenschaftlichen Arbeit, nicht die Lust entwickelt, Argumente zu finden, Theorien zu überprüfen und zu verwerfen und Alternativen zu suchen.

Vielen Dank an Katharina Maag Merki, die mir (als Politologin) die Arbeit in der Erziehungswissenschaft ebenso wie die Fallstudien in Kanada zugetraut hat, die mich durch Weiterbildungen und den Besuch von (inter-) nationalen Kongressen gefördert hat und stets Ansprechpartnerin für Fragen ist.

Danke an Lucien Criblez für treffend pointierte Nachfragen und Danke an André Kaiser und Ingo Rohlfing, die mich bei der Konzeption meiner Fallstudien unterstützt haben.

Außerdem möchte ich mich bei meiner Kollegin Britta Oerke für ihr differenziertes Feedback und ihre Vorliebe für leicht verständliche Sätze bedanken. Romanartige Artikel ließen sich so verhindern.

Danke auch besonders an die Stiftung für Kanada-Studien – ohne ihr Stipendium für die Kanadastudie hätte ich mich nicht für den wissenschaftlichen Weg entschieden, der mir nun so viel Freude bereitet. Danke auch an den DAAD, der mir ebenfalls finanziell mein Lieblingsprojekt, die best-case Studie zu Kanada, ermöglicht hat.

Und schliesslich ganz besonderen herzlichen Dank an die Ministerien, Schulen, LehrerInnen und SchülerInnen für die Beteiligung an den Projekten, und die Mitarbeit bei Umfragen und Interviews! Ohne diese Bereitschaft wäre ein praxisnahes Forschen und Analysieren nicht möglich gewesen.

## 1 Einleitung

Bildungssysteme sollen kulturelle Teilhabe ermöglichen, berufsrelevante Fähigkeiten vermitteln, soziale Positionen in der Gesellschaft durch die Auswahl nach Leistung verteilen und eine soziale Identität sowie politische Teilhabe gewährleisten (vgl. Fend, 2009, S.51). Seit Veröffentlichung der PISA Ergebnisse durch die OECD (u.a. 2007, 2010) ist jedoch das Bildungssystem mit dem Vorwurf konfrontiert, diese Aufgaben nicht adäquat zu bewältigen. Die Bildungsforschung fragt daher nach Faktoren, die die Leistung der Bildungssysteme verbessern können. Ansätze zur Verbesserung werden u.a. auf der Ebene des Unterrichts, in der Professionalisierung der Lehrpersonen oder aber auch in Reformen der Steuerung und Governance des Bildungssystems gesucht.

In der vorliegenden Arbeit wird angenommen, dass Governance-Strukturen einen wichtigen Beitrag zur Leistungsbereitstellung von Bildungsangeboten leisten. Dies bedeutet nicht, dass Governance-Strukturen Handlungen zu steuern vermögen, sondern lediglich, dass sie versuchen, Handlungsweisen einzuhegen (vgl. u.a. Mayntz, 2004; Altrichter & Maag Merki, 2010). Dies kann direkte wie indirekte, erwünschte wie nicht erwünschte Effekte hervorbringen (vgl. u.a. Altrichter & Maag Merki, 2010; Braun, 2008). In der vorliegenden Arbeit wird daher aus der Governance-Perspektive<sup>1</sup> gefragt, inwiefern bestimmte Koordinationsformen zwischen Akteuren erwünschte wie auch unerwünschte Effekte im Bildungssystem hinsichtlich der angestrebten Ziele hervorruufen.

Dieser Ansatz ist nicht nur theoretisch, sondern auch gesellschaftlich verankert: So zeugen zahlreiche Initiativen weltweit von der Idee, durch eine Reform der Strukturen des Bildungssystems die Ergebnisse der SchülerInnenleistungen beeinflussen zu können. Allein im deutschsprachigen Raum finden sich Beispiele zu umfassenden Reformen der Governance im Bildungsbereich: So beteiligen sich mehr als zwei Drittel der Schweizer Kantone an dem Projekt HarmoS<sup>2</sup>, bei welchem Strukturen der Bildungssysteme zur Qualitätssicherung vereinheitlicht werden. In fast allen deutschen Bundesländern wurden zentrale Prüfungen eingeführt (Ausnahme: Rheinland-Pfalz) (Kühn, 2012). In Österreich wurden u.a. Elemente der Schulautonomie implementiert (Altrichter & Heinrich, 2007). Auch von den BürgerInnen werden Veränderungen der Governance-Strukturen als Ansatz das Bildungssystem zu verbessern, teilweise begrüßt, wie z.B. eine Studie zur Einführung zentraler Prüfungen in Deutschland zeigt (Tillmann, Dederling, Kneuper, Kuhlmann & Nessel, 2008, S. 388).

Empirische Studien zur Wirkung von Governance-Modellen zeigen allerdings, dass bisher nicht klar ist, welche Modelle im Bildungssystem erfolgversprechend sind.

---

<sup>1</sup> Die Governance-Forschung erforscht „das Zustandekommen sozialer Ordnung („Regelungsstruktur“) und sozialer Leistungen („Leistungsstruktur“) unter der Perspektive der Handlungskoordination zwischen verschiedenen sozialen Akteuren in komplexen Mehrebenensystemen“ (Altrichter & Helm, 2011, S. 28), vgl. Kapitel 2.1.

<sup>2</sup> Vgl. <http://www.edk.ch/dyn/11659.php>, Stand: 18.05.2013

So resümieren Altrichter und Helm (2011, S. 31), dass weder bürokratische Modelle<sup>3</sup>, noch New Public Management Konzepte (NPM)<sup>4</sup> überlegen sind, wenn man die Leistungen der Bildungssysteme vergleicht.

Studien zu einzelnen Instrumenten dieser Modelle, wie zu Schulwahlfreiheit oder zu Wettbewerb zwischen Schulen, liegen vor (vgl. für einen Überblick alleine zu „Risiken Neuer Steuerung“ Bellmann & Weiß, 2009). Diese Studien weisen jedoch oft kein längsschnittliches Forschungsdesign auf. Da sich zudem die Bildungssysteme in ihren Gesamtarrangements stark unterscheiden, sind Übertragungen von Ergebnissen oftmals schwierig. Insgesamt liegt demnach wenig Wissen zu Governance-Strukturen unter Berücksichtigung der Mehrebenenstruktur vor und je nach Steuerungsinstrument und Konfiguration des Bildungssystems bedarf es weiterer Forschung dazu, wie das entsprechende Instrument wirkt.

Die vorliegende Arbeit knüpft an diese Desiderata mit empirisch ausgerichteten Studien an. Dabei wird einmal „rückwärtsblickend“ (vgl. Scharpf, 2000, S. 57) nach Gelingensbedingungen von Governance in einem best-case in Alberta, Kanada, (Studie A) und einmal „vorwärtsblickend“ (vgl. Scharpf, 2000, S. 59) nach nicht intendierten Effekten eines Steuerungsinstrumentes, des Zentralabiturs in Hessen und Bremen, in Deutschland, (Studien B, C und D), gefragt. Die Titel der Studien lauten wie folgt:

- Studie A (best-case Studie): Zwischen Empowerment und Kontrolle. Die praktische Umsetzung des New Public Management und Professioneller Lerngemeinschaften in Kanada. Eine Fallstudie (Jäger, 2014)
- Studie B: State-wide, low-stakes tests and a teaching-to-the-test effect? An analysis of data from two German Bundesländern (Jäger, Maag Merki, Oerke & Holmeier 2012)
- Studie C: Herausforderung Zentralabitur: Unterrichtsinhalte an Prüfungsthemen anpassen, ohne dabei die Interessen der Schüler/-innen zu vernachlässigen (Jäger, 2012a)
- Studie D: Schulklima, Selbstwirksamkeit und Arbeitszufriedenheit aus Sicht der Lehrpersonen und Schüler/-innen in Hessen und Bremen (Jäger, 2012b)

Die best-case Analyse in Kanada (Studie A) wurde durchgeführt, da die Wirkung von Governance-Strukturen umstritten ist und sich kein „ideales“ Governance-Modell abzeichnet. Die Studie eruiert, welche Kombination von Instrumenten und Mechanismen in der spezifischen Situation im untersuchten Fall in Alberta eingesetzt wird und dort zum Untersuchungszeitpunkt Erfolg erzielt. Laut Literatur sind in Alberta New Public Management Instrumente wie freie Schulwahl und Schulautonomie sowie Professionelle Lerngemeinschaften installiert (vgl. Evans, 1997; Bonsen & Rolff, 2006; OECD, 2004; Rolff, 2001; Tarnoczi, 2006). Wie dies umgesetzt wird, wird in der Fallstudie untersucht. Die Ergebnisse können zwar nicht verallgemeinert werden, sie können aber

---

<sup>3</sup> Charakteristika bürokratischer Modelle sind u.a. hierarchische Beziehungen und Inputsteuerung (vgl. Altrichter & Helm, 2011)

<sup>4</sup> Unter New Public Management fallen Reformen der öffentlichen Verwaltung, die Output-basierte Instrumente der Steuerung beinhalten (Schedler & Proeller, 2006). Charakteristika von New Public Management sind u.a. zentrale Vorgaben und dezentrale Ressourcenverwaltung, vgl. Kapitel 6.1.



Anhaltspunkte liefern, um erfolgreiche Lösungsmuster zu identifizieren, zu ordnen und für andere Situationen, Bereiche und Zeitpunkte fruchtbar zu machen.

Die drei Studien (B, C und D) zur Auswirkung des Steuerungsinstrumentes Zentralabitur untersuchen die Wirkung der Implementierung dieses Instrumentes hinsichtlich relevanter Bereiche für die Schulqualität: So hinsichtlich der Wirkung zentraler Prüfungen auf die Arbeitsunzufriedenheit, die Wahrnehmung der individuellen und kollektiven Selbstwirksamkeit, das Schulklima und die Themenvarianz im Unterricht (Teaching to the test<sup>5</sup>). Es wird u.a. argumentiert, dass das Instrument Zentralabitur als ein ex post Kopplungsmechanismus eingeordnet werden kann und hier u.a. die Gefahr eines „Übergeneralisierungseffektes“ birgt (vgl. Motivation-Crowding Theorie, u.a. Frey & Jegen, 2002). Das heißt, dass die durch das Zentralabitur vorgegebenen Regeln z.B. im Bereich der Themenvarianz im Unterricht zu eng ausgelegt werden und es zu einem Teaching to the test kommt. Anhand empirischer Daten wird die Auswirkung der Einführung zentraler Prüfungen auf die verschiedenen Bereiche der Schulqualität quer- und längsschnittlich untersucht. Damit befassen sich die drei Studien mit der Output-Seite des Bildungssystems und hier besonders mit nicht intendierten Effekten (z.B. Teaching to the test).

#### *Originalität der einzelnen Studien*

Die Originalität der Studie A liegt in ihrem Fokus auf das gesamte Bildungssystem und der detaillierten Analyse der Koordinationsmechanismen zwischen den Akteuren. Alle für die Schule relevanten Akteure werden in die Analyse einbezogen und nach ihren Beziehungen zueinander befragt. Die Fallstudie verbindet darüber hinaus qualitative wie quantitative Erhebungsinstrumente. Sie wurde in Eigenregie geplant, durchgeführt und ausgewertet<sup>6</sup> und knüpft an die Forschungslücke an, die Altrichter & Helm (2011), Berkemeyer (2010) und Thillmann (2011, S. 75) konstatieren, nämlich, dass Wissen zu erfolgreichen Steuerungskonfigurationen bzw. zu Merkmalskombinationen fehlt, in denen das Mehrebenensystem betrachtet wird.

Die Originalität an den Auswertungen zum Zentralabitur in Deutschland ist, dass es sich um Längsschnittstudien handelt, die zudem im Bundesland Bremen den Wechsel von einem dezentralen zu einem zentralen Prüfsystem verfolgen konnten. Die Leistung meiner Beiträge im Rahmen des von Prof.in Katharina Maag Merki und Prof. Eckhardt Klieme geleiteten Projektes ist die theoretische Herleitung nicht intendierter Effekte aufgrund der Einführung zentraler Prüfungen unter low stakes Bedingungen<sup>7</sup> (vgl. Kapi-

---

<sup>5</sup> Themenvarianz wird hier u.a. verstanden als das Ausmaß, in dem die Interessen der Schüler/-innen und aktuelle Themen im Sachfach bei der inhaltlichen Auswahl der Unterrichtsthemen berücksichtigt werden. Teaching to the test ist hier definiert als das Gegenteil. Also, wenn in diesem Sinne keine Themenvarianz im Unterricht verwirklicht wird.

<sup>6</sup> Herzlichen Dank an dieser Stelle für die Förderung dieser Studie an die Stiftung für Kanada Studien und den Deutschen Akademischen Austauschdienst!

<sup>7</sup> Als „high stakes“ Systeme werden jene Bildungssysteme bezeichnet, in denen aufgrund von Evaluationsergebnissen oder vergleichenden Prüfungen Boni bzw. Sanktionen für Lehrkräfte, Schulleitungen oder gesamte Schulen vergeben werden (vgl. Madaus, 1988, S. 87, zitiert nach Andrews, 2004). Low-stakes Systeme haben keine solchen Sanktionen implementiert.

tel 7). Bei den empirischen Auswertungen wurde die Mehrebenenstruktur des Bildungssystems durch Mehrebenenanalysen mittels HLM 6.06 (Raudenbush, Bryk und Congdon, 2004) berücksichtigt.

#### *Gemeinsamer Hintergrund der Studien: die Governance-Perspektive*

Die vier Studien basieren auf der Annahme, dass Organisationen Unterschiede erzeugen (vgl. March & Olsen, 1984, S.747). March & Olsen (1984, S. 747) fügen kritisch hinzu: „The new institutionalism is an empirically based prejudice“. Inwiefern es zutrifft oder widerlegt wird, muss die Zukunft zeigen. Zudem wird angenommen, dass absichtsvolle Interventionen Ergebnisse u.a. der Bildungssysteme – trotz möglicher Interessenskonflikte zwischen Prinzipal und Agent in beabsichtigter Weise beeinflussen können (vgl. Prinzipal-Agent Theorie; Stewardshiptheorie; Döhler, 2008; Manow, 2008; dargelegt in Kapitel 3.1). In den einzelnen Analysen stehen argumentativ Kopplungen und ihre - Mechanismen als Teil der Regelungsstruktur von Governance im Vordergrund. Kopplungen bezeichnen all das, was Personen oder Institutionen verbindet (Lasky, Datnow & Stringfield, 2005).

So fokussiert die best-case Studie A Kopplungsmechanismen zwischen den Akteuren im Mehrebenensystem, um die Umsetzung der Governance-Strukturen zu untersuchen. Die Modelle des New Public Management und der Professionellen Lerngemeinschaften (PLG)<sup>8</sup> werden dabei als Grundlage genutzt, da beide de jure in Alberta implementiert sind. Die Arbeit untersucht jedoch mittels einer Netzwerkanalyse systematisch alle Beziehungen und nimmt damit zusätzlich explorativ alle Kopplungen zwischen Akteuren in den Blick.

In den Studien zum Zentralabitur in Hessen und Bremen (Studien B, C und D) werden Kopplungen reflektiert, da das Zentralabitur eine zweifache Reform der Kopplungsmechanismen beinhaltet: Zum einen wird die Arbeit der Lehrenden ex post kontrolliert, also nach der Leistungserbringung, und damit mit den Vorgaben des Ministeriums gekoppelt. Zweitens werden die Lehrpersonen von der Prüfung selbst entkoppelt, da die Prüfung nun außerhalb der Schule seitens des Ministeriums formuliert wird und nicht mehr durch die Lehrpersonen selbst. In den Studien wird argumentiert, dass diese spezielle Konfiguration der Kopplungsmechanismen zu nicht intendierten Effekten führt.

In der vorliegenden Synopse, die die Studien rahmt, werden daher Kopplungen und ihre Mechanismen fokussiert, um die folgenden übergeordneten Forschungsfragen zu bearbeiten:

- I. Welche Kopplungsmechanismen können dazu beitragen, dass in Bildungssystemen Ziele umgesetzt werden, das heißt, die erwünschte Leistung erbracht wird?
- II. Wie wirken sich Reformen der Kopplungsmechanismen aus?

Anhand der Literatur werden Kopplungsmechanismen differenziert betrachtet, indem u.a. ihre Multiplexität herausgestellt wird. Der Begriff Multiplexität ist aus der Netz-

---

<sup>8</sup> Zu Elementen der Professionellen Lerngemeinschaften gehören horizontale Strukturen, Professionalisierung und Kollegialität (vgl. Marzano, 2003, S. 6, zit. nach Alberta Education, 2006).

werkforschung entlehnt und bezeichnet den Sachverhalt, dass Personen auf verschiedenen Dimensionen miteinander verbunden sein können (unterschiedliche Beziehungstypen: z.B. familiär oder geschäftlich) und damit Netzwerke multiplex strukturiert sind (vgl. Jansen, 2006, S. 108f). Zudem wird in dieser Arbeit argumentiert, dass die Kombination unterschiedlicher Kopplungsmechanismen im Bildungssystem aufgrund dessen Eigenlogik sinnvoll erscheint, wenn eine absichtsvolle Intervention gelingen soll.

Die oben genannten Fragen werden innerhalb der Arbeit theoretisch problematisiert und diskutiert. Zudem liefern die Studien A, B, C und D Ergebnisse zu Teilaspekten dieser Fragestellungen. Sie bearbeiten folgende Teilfragen:

1. Mittels welcher Kopplungsmechanismen wird in einem best-case gesteuert?
2. Erzeugen ex post Kopplungsmechanismen nicht intendierte, unerwünschte Effekte?

#### *Aufbau der Synopse*

In dieser Arbeit wird folgendermassen argumentiert: Zuerst (Kapitel 2) werden die Begrifflichkeiten Governance, Steuerung, Kopplungen, Kopplungsmechanismen und (nicht) intendierte Effekte in ihrer hier vorliegenden Verwendung geklärt. Im Anschluss (Kapitel 3) werden typische Probleme der Steuerung und Governance, nämlich nicht intendierte Effekte und ihre Ursachen, erörtert. Zudem wird eruiert, inwiefern Information, Koordination und bestimmte Kopplungsmechanismen das Auftreten nicht intendierter Effekte vermindern können. Nach diesen theoretischen Ausführungen werden in Kapitel 4 Charakteristika des Bildungssystems wie die Technologievagheit aufgezeigt und deren Implikationen für Kopplungsmechanismen und Governance abgeleitet.

Kapitel 5 fasst die Argumentation der vorherigen Artikel zusammen und leitet daraus die übergeordneten Forschungsfragen sowie die Teilfragen für die präsentierten Studien ab. In Kapitel 6 und 7 werden dann Erwartungen für die Teilstudien formuliert und die Ergebnisse der Studien präsentiert und diskutiert. Dabei richten die Teilstudien den Blick einmal auf Gelingensbedingungen der Governance und eruieren dort, wie Kopplungsmechanismen funktionieren. Zum Zweiten wird in den Studien die Wirkung einer Reform der Kopplungsmechanismen, dem Zentralabitur, überprüft. Die Diskussion (Kapitel 8) führt die Ergebnisse der Teilstudien und der theoretischen Problematisierung der Synopse zusammen.

## **2 Governance, Steuerung, Kopplungen, Kopplungsmechanismen und (nicht) intendierte Effekte**

Die hier präsentierten Studien A, B, C und D basieren auf der Annahme, dass Governance-Strukturen sich auf die Bereitstellung von Leistungen auswirken. Sie untersuchen daher einmal rückwärtsblickend, welche Faktoren positiv wirken könnten (Studie A) und vorwärtsblickend, welche nicht intendierten Effekte Interventionen wie das Instrument der zentralen Abschlussprüfung am Gymnasium hervorrufen (vgl. Scharpf, 2000, S. 57-59). In den folgenden Abschnitten wird nun definiert, was unter Governance, Steuerung, Kopplungen, Kopplungsmechanismen und (nicht) intendierten Effekten verstanden wird. Governance wie Steuerung beschäftigen sich mit der „absichtsvollen Regelung“ (vgl. Mayntz, 2008, S. 55), doch diese ist kaum ohne nicht intendierte Effekte denkbar (vgl. Braun, 2008). Um erfolgreiche Governance zu messen, muss demnach neben der Umsetzung der intendierten Ziele auch immer mitberücksichtigt werden, ob eine Intervention negative, unerwünschte Effekte hat vermeiden können (Braun, 2008, S.104). Daher wird weiter dargelegt, welche Ursachen für nicht intendierte Effekte in der Literatur diskutiert werden und es wird herausgestellt, inwiefern Kopplungen bzw. ihre Mechanismen mit dem Auftreten nicht intendierter Effekte zusammenhängen.

### **2.1 Governance und Steuerung**

Die Governance-Forschung erforscht „das Zustandekommen sozialer Ordnung („Regelungsstruktur“) und sozialer Leistungen („Leistungsstruktur“) unter der Perspektive der Handlungskoordination zwischen verschiedenen sozialen Akteuren in komplexen Mehrebenensystemen“ (Altrichter & Helm, 2011, S. 28). Dies bedeutet, dass die Educational Governance-Forschung unter Berücksichtigung u.a. der BildungspolitikerInnen, der Verwaltung, der Schulleitung, der Lehrpersonen und SchülerInnen Gesetze, Vorgaben, Regelungen (Regelungsstrukturen) sowie ihre Umsetzung in der Praxis, die Bereitstellung von Bildungsangeboten, untersucht. „Die Grundfrage der Governanceforschung [...] lautet: Erfüllen die Regelungsstrukturen ihre Funktion? Stellen sie tatsächlich sicher, dass die Organisation ihre Leistungen erfüllt – oder nicht?“ (Langer, 2009c, S. 510).

Diese Auslegung des Governance-Begriffs grenzt sich von einem normativen Verständnis von Governance ab und betont den Charakter des Ansatzes als Analyseinstrument (vgl. Langer & Hüther, 2009; Mayntz, 2004; Schimank, 2007). So werden unter der analytischen Perspektive alle Arten der Koordination und ihr Verhältnis zueinander rekonstruiert (Langer & Hüther, 2009, S. 490). VertreterInnen der normativen Verwendungsweise des Begriffs bezeichnen mit Governance dagegen *eine* bestimmte Form des Regierens: "For one thing, 'governance' is now often used to indicate a new mode of governing that is distinct from the hierarchical control model, a more cooperative mode where state and non-state actores participate in mixed net-

works“ (Enders, 2002, S. 8-9, zitiert nach Langer & Hüther, 2009; vgl. u.a. Wald & Jansen, 2007; World Bank<sup>9</sup>).

Der hier verwendete Governance-Begriff richtet somit sein Augenmerk auf alle möglichen Regelungsstrukturen (Staat, Markt, Gemeinschaft) und Modi der Handlungskoordination – Hierarchie, Konkurrenz und Verhandlung/ Solidarität – und möchte die Kombination dieser beschreiben und analysieren (vgl. u.a. Langer & Hüther, 2009, S. 476).

Um anschaulicher zu beschreiben, was mit Governance gemeint ist, grenzt Schuppert (2008, S. 22f) den Begriff Governance zu Management, Government und Steuerung ab: So steht im Vergleich zum Management bei der Governance nicht die „Binnensicht auf die Verwaltung“ im Vordergrund, sondern das „Verhältnis von staatlichen und nicht-staatlichen Akteuren“ (Schuppert, 2008, S. 22). Dies wird deutlich an den „von Jann als governancetypisch reklamierten Begriffen“: u.a. Koordination öffentlicher und gesellschaftlicher Akteure, Koordination verschiedener Steuerungsformen, neue Aufgabenteilung, Koproduktion und Selbstregelung (vgl. Schuppert, 2008, S. 22).

Im Vergleich zu dem Ansatz des Government fokussiert der Begriff Governance – wie bereits erwähnt – nicht nur den Staat und die Hierarchie als wichtigsten Handlungsmodus, sondern u.a. die institutionelle Struktur, Netzwerke, Verhandlung als Handlungsmodus zwischen staatlichen und/oder gesellschaftlichen Akteuren (vgl. Schuppert, 2008, S. 22f). Mayntz (2008, S. 55) führt ebenfalls aus, dass der Begriff Governance die Aufmerksamkeit auf „das Zusammenspiel, die Kombinierbarkeit verschiedener Governance-Formen, ihre gegenseitige Ergänzung oder gegenseitige Beeinträchtigung“ lenkt (Mayntz, 2008, S. 55). Durch die Governance-Forschung werden somit Mischformen der Handlungskoordination sichtbar (vgl. Schuppert, 2008, S. 24).

Als drittes Begriffspaar unterscheidet Schuppert (2008) Steuerung und Governance. Hier führt er in Anlehnung an Mayntz aus, dass Steuerung und Governance in Ergänzung zueinander betrachtet werden sollten (vgl. Schuppert, 2008, S. 23). In der Steuerungstheorie wird untersucht, unter welchen Bedingungen eine intentionale Beeinflussung der Handlung von Person B durch Person A gelingt (vgl. Mayntz, 2004). Der Begriff der Steuerung wird hier nicht als ein „top-down-Prozess“ verstanden, sondern als ein Begriff, der alle Formen der Beeinflussung zwischen Personen in den Blick nimmt. In diesem Sinne ist er komplementär zu dem Begriff der Governance. Governance bezeichnet das „Gesamt aller auf einer bestimmten territorialen Ebene nebeneinander bestehenden Formen der absichtsvollen Regelung gesellschaftlicher Sachverhalte“ (Mayntz, 2008, S. 55). Die Konzepte der Steuerung und Governance beschäftigen sich somit beide mit der „intentionalen Beeinflussung“ bzw. „absichtsvollen Regelung“ zwischen Akteuren bzw. Institutionen, lediglich die Analyseperspektive ist unterschiedlich:

---

<sup>9</sup> Vgl.

<http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/MENAEXT/EXTMNAREGTOPGOVERNANCE/0,,contentMDK:20513159~pagePK:34004173~piPK:34003707~theSitePK:497024,00.html>; Letzter Zugriff: 13.08.2013

„Die mit den Begriffen ‚Steuerung‘ bzw. ‚Governance‘ verbundenen Paradigmen lenken die Aufmerksamkeit auf verschiedene Aspekte der gleichen Wirklichkeit (Mayntz, 2005). In der Steuerungstheorie stehen (korporative) Akteure und ihr Handeln im Zentrum des Interesses. Der Ansatz der Governance-Theorie ist dagegen strukturzentriert; hier stehen Institutionen, steht die Regelungsstruktur und ihre Leistungsfähigkeit im Zentrum des Interesses“ (Schuppert, 2008, S. 23).

Beide Begriffe lenken also die Aufmerksamkeit auf unterschiedliche Teilaspekte und bieten damit unterschiedliche Vorteile: Der Vorteil des Begriffs Steuerung gegenüber dem Governance-Begriff liegt darin, dass er zwischen Steuerungssubjekt und -objekt unterscheidet (Mayntz, 2008, S. 44; Offe, 2008, S. 66). So kann Verantwortung für Handlungen und deren Folgen Akteuren zugeschrieben werden.<sup>10</sup> Der Vorteil des Governance-Begriffs gegenüber dem Steuerungsbegriff ist, dass er die Perspektive auf Strukturen und deren Leistungsfähigkeit ausweitet. So kann die Komplexität der Realität adäquater erfasst werden. Die Begriffe werden im Weiteren als sich ergänzend verstanden.

Mit Governance und Steuerung ist die Frage verknüpft, ob eine erbrachte Leistung den gesetzten Zielen entspricht. Dies impliziert zwei Probleme, die in der Prinzipal-Agenten Theorie<sup>11</sup> als „hidden information“ und „hidden action“ bezeichnet werden (Strøm, 2000, S. 270f). Diese Begriffe beziehen sich auf Delegationsbeziehungen, sie können jedoch auch im Rahmen der Governance-Analyse als Heuristik genutzt werden. „Hidden information“ bezeichnet in diesem Rahmen das Problem, dass zumeist zum Zeitpunkt von Entscheidungen zu wenig Wissen über die Folgen der Entscheidung vorliegt und dass die Person, die beeinflusst, zumeist über weniger Wissen zur konkreten Umsetzung verfügt als die zu beeinflussende Person. „Hidden action“ benennt hier das Problem, dass die Akteure, die eine Entscheidung umsetzen, immer auch über Möglichkeiten verfügen, von Vorgaben abzuweichen. Beides resultiert womöglich darin, dass nicht die Ziele erreicht werden, die durch eine Intervention beabsichtigt sind.

## 2.2 *Kopplungen und Kopplungsmechanismen*

Kopplungen bilden einen Teil der Regelungsstrukturen der Governance, sie können formalisiert oder auch spontan erfolgen und umfassen alle Arten der Verknüpfung zwischen Individuen, Systemen bzw. Institutionen. Kopplungen zu analysieren, macht Handlungsspielräume und -optionen sichtbar (vgl. Lasky, Datnow & Stringfield, 2005, S. 241). Denn der Begriff Kopplungen bezeichnet auch Beziehungen, die noch nicht aktiviert wurden, aber prinzipiell möglich sind. Nicht aktivierte Beziehungen sind z.B. die Beziehungen zwischen Teilnehmenden einer Weiterbildung. Sie teilen sich vorerst nur einen Raum, es existiert aber die Möglichkeit sich auszutauschen. Der Begriff der

---

<sup>10</sup> Der Begriff der Begriff Governance wäre dafür nicht hilfreich, verwischt er doch nicht weniger als sieben Distinktionen (Offe, 2008, S. 66): Subjekt/Objekt, Staatlich/Gesellschaftlich, politisch/wirtschaftlich, Struktur/ Prozesse, innen-/außenpolitisch/international, beobachtbare Vorgänge/ soziale Normen, Entwicklungs-/Schwellenländer/OECD-Staaten.

<sup>11</sup> Die Prinzipal-Agent Theorie beschäftigt sich mit Delegationsbeziehungen zwischen mindestens zwei Personen: dem Auftraggebenden (Prinzipal) und dem Auftragnehmenden (Agent) (vgl. für einen Überblick Gilardi & Braun, 2002).

Kopplungen, wie er hier in Anlehnung an die Netzwerktheorie verwendet wird, bezeichnet damit alle Gelegenheiten der Koordination (Jansen, 2006).

Glassman (1973) definiert die Stärke der Kopplung über die Schnittmenge, die zwei Systeme miteinander teilen. Weick (1976/ 2009, S. 89) beschreibt die Stärke der Kopplung nach Glassman folgendermassen: „Nimmt man Schulleiter-stellvertretenden Schulleiter-Schulinspektor als ein System an und Lehrer-Klassenraum-Schüler-Eltern-Curriculum als ein anderes, dann kann der Schulleiter als lose gekoppelt mit dem Lehrer verstanden werden, wenn nur wenige Variablen von Bedeutung in der Welt des Lehrers auch in der Welt des Schulleiters eine wichtige Rolle spielen und/oder wenn die gemeinsam geteilten Variablen unbedeutend sind im Vergleich zu anderen Variablen in den jeweiligen Welten.“ Weick (1976/ 2009, S. 92) definiert in Anlehnung an Glassman (1976) den Grad der Kopplung als die mit der Handlung verbundenen Voraussetzungen der Handlung und schreibt: „je länger die Kette von Voraussetzungen, desto enger die Kopplung“.

Allerdings ist bisher wenig erforscht, welche Art der Kopplung wie wirkt. Die Zusammenhänge auf und zwischen den Ebenen im Bildungssystem, „also die verschiedenen Kopplungen, sind bislang weder hinreichend beschrieben noch in Ansätzen erforscht (Berkemeyer, 2010, S.144). In der Literatur wird oftmals die „besondere“ Situation der „losen Kopplung“ der Akteure (Weick, 1976/ 2009), insbesondere der Lehrpersonen, hervorgehoben (vgl. u.a. Drepper, 2010).

Unter loser Kopplung wird verstanden, dass u.a. die Lehrperson im Unterricht spontan reagieren muss und es nicht möglich ist, hier situationsspezifisch durch Vorgaben die Handlungen im Detail seitens der Schulleitung oder des Ministeriums zu bestimmen. Diese fehlende Kontrolle des Agenten (Auftragnehmenden) – um es im Jargon der Prinzipal-Agent-Theorie auszudrücken – mag seitens des Prinzipals (Auftraggebenden) als problematisch oder aber als Chance eingeordnet werden. Denn durch den Freiraum in der losen Kopplung können die Lehrpersonen einerseits von Vorgaben abweichen, andererseits können sie im Idealfall zielgenau ihre Reaktion in der spezifischen Situation anpassen.

Schliesslich können Kopplungen von Kopplungsmechanismen unterschieden werden: Kopplungsmechanismen bezeichnen all jene Kopplungen, die wiederkehren und bestimmte Wirkungen auslösen: So werden als Mechanismen „wiederkehrende Prozesse, die bestimmte Ursachen mit bestimmten Wirkungen verbinden“, bezeichnet (Mayntz, 2002, S. 24).

Kopplungen wie -mechanismen sind insofern im Rahmen der Governance-Forschung relevant, da sie den Blick auf die potentiellen sowie die realisierten Beziehungen lenken und zudem zusätzliche Gestaltungsspielräume für die Politik oder die Organisation von Institutionen bieten. Möglichkeiten der Begegnung können institutionell geschaffen werden. Gelegenheiten des Austauschs können durch das Einrichten von (Zeit-) räumen in Organisationen verankert werden. Kopplungen wie Kopplungsmechanismen können demnach wichtige Bausteine zur Generierung adäquater Ziele und zur Durchsetzung und Umsetzung dieser darstellen – allerdings nur unter der Voraussetzung, dass die Möglichkeiten und Gelegenheiten auch qualitativ hochwertig genutzt werden (vgl. Lasky, Datnow & Stringfield, 2005). Im Folgenden wird der Begriff

Kopplungsmechanismen differenziert erörtert. Dies ist hilfreich, da angenommen wird, dass Kopplungsmechanismen unterschiedlich Steuerung bzw. Governance und damit das Auftreten intendierter und nicht intendierter Effekte beeinflussen.

Manow (2005, S. 246) definiert Kopplungsmechanismen zeitlich als ex ante, ex ongoing und ex post Kontrolltechniken, also als Techniken der Verbindung oder Kontrolle, die vor, während oder nach einer Leistungserbringung erfolgen. Ex ante Kontrolle vollzieht sich z.B. durch Ausbildungsgrundlagen, das Vertragsdesign zwischen Prinzipal und Agent, und durch die KandidatInnenauswahl. Ex ongoing Kontrolle ist beispielsweise Monitoring während der Prozesse und ein System aus „checks and balances“ (Strøm, 2000, S. 271 - 275). Ex post Kontrolle sind z.B. Belohnungen und Sanktionen für erbrachte Leistungen. Das Zentralabitur ist demnach eine ex post Kontrolle der Leistungen der Lehrpersonen und SchülerInnen während der Oberstufenzeit.

Glassman (1973) wie Weick (1976 / 2009) unterscheiden die Stärke der Kopplungen anhand der geteilten Schnittmenge der Akteure bzw. Systeme. Analog kann man annehmen, dass auch die Stärke der Kopplungsmechanismen von der gemeinsamen Schnittmenge abhängt. Kopplungen – wie Kopplungsmechanismen – sind dann jedoch nicht nur auf Kontrollfunktionen begrenzt, sondern bezeichnen auch z.B. ideologische Übereinstimmungen zwischen Akteuren. Kopplungen werden dementsprechend auch als Verbindungen allgemein definiert, die sich wie Lasky, Datnow und Stringfield (2005, S. 241) aufzeigen, in fünf Arten von Kopplung unterscheiden: formal (hergestellt durch ein offizielles Mandat), informell (Beziehungen z.B. auch nur über Email und Telefon), strukturell (z.B. durch Finanzierung der Schulen durch den Staat), relationale Kopplungen (z.B. zwischen KollegInnen) und schliesslich weltanschauliche Kopplungen (z.B. geteilte Ideen und Ziele) (vgl. Lasky, Datnow & Stringfield, 2005, S. 241).

Diese Typologie zeigt, dass Kopplungen multiplex<sup>12</sup> betrachtet werden müssen. So ist das Zentralabitur ein formaler ex post Kopplungsmechanismus zwischen Ministerium und Lehrperson, es könnte aber zusätzlich durch z.B. Fortbildungen zu den zentralen Prüfungen relational gekoppelt werden. Die Typologie ist allerdings nicht trennscharf. So sind formale und informelle Kopplungen zwei Pole zwischen denen Beziehungen realisiert werden können. Relationale Kopplungen bilden eine eigene Form der Kopplung, obwohl sie ebenfalls auf diesem Kontinuum zwischen formal und informell angesiedelt werden könnten. Das gleiche gilt für strukturelle Kopplungen: Sie zeigen, dass formale Kopplungen durch Mandat, aber auch durch finanzielle Abhängigkeiten aufrechterhalten werden können. Zudem ist fraglich, wo die losen Kopplungen (Weick, 1976/ 2009) eingeordnet werden. Auch sie könnten – im Gegensatz zu den informellen Kopplungen – als Gegenpol zu den formalen Kopplungen definiert werden.

An dieser Stelle ist nicht der Platz, um diese Typologie umfassend zu diskutieren oder zu ergänzen. Die Typologie wird hier als hilfreiche Arbeitshypothese verstanden, die in weiteren Studien weiterentwickelt werden muss. An dieser Stelle ist wichtig, dass Kopplungen multiplex auftreten: Z.B. können zwei Akteure weltanschaulich *und* formal

---

<sup>12</sup> Vgl. Fußnote 10: Multiplexität bezeichnet den Sachverhalt, dass Personen auf verschiedenen Dimensionen miteinander verbunden sein können (unterschiedliche Beziehungstypen: z.B. familiär oder geschäftlich) (vgl. Jansen, 2006, S. 108f).



gekoppelt sein, sie können informell *und* relational gekoppelt sein, aber auch nur strukturell.

Eine weitere Unterscheidung von Kopplungen findet sich bei Richter (2012, S. 99). Er betrachtet Kopplungen von der zu erbringenden Leistung her und differenziert, ob die Leistungserbringung selbst inhaltlich oder zeitlich begrenzt ist. Der Unterschied zu Manow (2008) liegt darin, dass nicht, der Zeitpunkt der Kontrolle als Kriterium für die Kopplung herangezogen wird, sondern die Dauer und der Umfang der Leistungserbringung. So weist Richters Unterscheidung darauf hin, dass Kopplungen auch nur für einen bestimmten Teil der Leistungserbringung gelten können.

Tabelle 1: Kopplungsmechanismen

Kopplungsmechanismen zur Sicherstellung der Leistungserbringung				
Wann?	Ex ante	Ex ongoing	Ex post	
Wie?	<ul style="list-style-type: none"> <li>• Formal</li> <li>• Informell</li> <li>• Strukturell</li> <li>• Relational</li> <li>• Weltanschaulich</li> </ul>	<ul style="list-style-type: none"> <li>• Formal</li> <li>• Informell</li> <li>• Strukturell</li> <li>• Relational</li> <li>• Weltanschaulich</li> </ul>	<ul style="list-style-type: none"> <li>• Formal</li> <li>• Informell</li> <li>• Strukturell</li> <li>• Relational</li> <li>• Weltanschaulich</li> </ul>	
Was?	Leistungserbringung	Leistungserbringung	Leistungserbringung	
	<ul style="list-style-type: none"> <li>• Thematisch festgelegt</li> <li>• Temporär bestimmt</li> </ul>	<ul style="list-style-type: none"> <li>• Thematisch festgelegt</li> <li>• Temporär bestimmt</li> </ul>	<ul style="list-style-type: none"> <li>• Thematisch festgelegt</li> <li>• Temporär bestimmt</li> </ul>	

Quelle: Eigene Darstellung

Das Zentralabitur am Ende der Schulzeit kontrolliert z.B. die erbrachte Leistung der Lehrpersonen, welche thematisch wie zeitlich in der Oberstufe vermittelt werden soll, nicht aber die Leistung der gesamten Schulzeit. Insgesamt ist das Zentralabitur damit als ein formaler, thematisch und temporär begrenzter ex post Kopplungsmechanismus zwischen Lehrperson/ SchülerInnen und Ministerium einzuordnen.

Tabelle 1 zeigt die unterschiedlichen Typologien für Kopplungsmechanismen im Überblick. Zur Systematisierung dienen drei Fragen: (1) Wann findet die Kopplung statt (vor, während oder nach der Leistungserbringung)? (2) Wie wird gekoppelt? (3) Was wird gekoppelt?

Aus diesen verschiedenen Arten und Ausprägungen von Kopplungen lässt sich ableiten, dass Kopplungen multiplex betrachtet werden müssen (vgl. Tabelle 1). D.h. eine Situation darf nicht nur punktuell betrachtet werden, sondern z.B. die Situation der Lehrperson im Unterricht ist zwar teilweise von der Schulleitung entkoppelt, jedoch ist die Unterrichtssituation zeitlich begrenzt und diesem Prozess gehen ex ante Kopplungsmechanismen voraus und ex post Mechanismen folgen. Zudem können auch während des Unterrichtens, also während einer formal entkoppelten Zeitperiode, weltanschauliche Kopplungen greifen. Die Teilung von gemeinsamen Werten liefert somit den Rahmen für die Handlungsweisen im Unterricht. Die Lehrperson ist somit zwar formal im Unterricht (ex ongoing) von der Schulleitung, der Bildungsadministration und der Bildungspolitik entkoppelt, jedoch ist sie ex ante durch ihre Ausbildung gekop-

pelt sowie weltanschaulich im Unterricht durch Gesetze an die Ziele des Ministeriums und Vorgaben der Schulleitung gebunden (ex ongoing).

Nachdem nun im Folgenden die Begriffe intendierte und nicht intendierte Effekte erläutert werden, wird anschließend dargelegt, inwiefern Kopplungsmechanismen mit (nicht) intendierten Effekten in einem Zusammenhang stehen.

### 2.3 *Intendierte und nicht intendierte Effekte*

Steuerung wie Governance zielen auf eine „absichtsvolle Regelung“ (Mayntz, 2008). Das bedeutet, absichtsvolle Regelungen werden mit dem Ziel aufgestellt, intendierte Ziele umzusetzen. So gilt ein Effekt als beabsichtigt, wenn die Handlungsfolgen mit dem zuvor definierten Ziel übereinstimmen. „Nicht intendiert“ bedeutet analog, dass die Handlungsfolgen nicht dem Handlungsziel entsprechen (Bös, 2002, S. 7). Nicht intendierte Effekte können Nebeneffekte sein, d.h. sie können neben der erfolgreichen Umsetzung einer Intention auftreten, sie können aber auch den Erfolg einer Handlung untergraben oder ins Gegenteil kehren. Letzteres ist sozusagen der Gegenpol zum intendierten Effekt, da er die „kontraproduktiven Folgen“ bezeichnet (Bös, 2002, S.7). Betrachtet man aber nur die „gewöhnlichen“ nicht intendierten Effekte, so ist aufgrund der Unabsichtlichkeit noch nicht gesagt, ob die zu beobachteten Handlungsfolgen erwünscht oder unerwünscht sind. So schreibt Merton (1936, S. 895): „undesired effects are not always undesirable effects“.

Braun (2008, S. 105) unterscheidet drei Dimensionen nicht intendierter Effekte<sup>13</sup>: „vorhersehbar-unvorhersehbar“, „erwünscht-unerwünscht“ und „beherrschbar-unbeherrschbar“. Effekte, die nicht vorhersehbar und nicht erwünscht sind, sind die Effekte, die allgemeinsprachlich als nicht intendiert und als ein Steuerungsproblem bezeichnet werden. Positive Nebeneffekte, die man zwar nicht vorhergesehen hat, die aber erwünscht sind, werden dagegen billigend in Kauf genommen oder begrüßt (Braun, 2008, S. 105). Die dritte Dimension ist hinsichtlich der vorhersehbaren Effekte interessant: sie können beherrschbar und unbeherrschbar sein. Als unbeherrschbar gelten z.B. Situationen, in denen Akteure gegenseitig aufeinander reagieren und sich schon ein Muster eingespielt hat (vgl. Braun, 2008, S. 112 – 120). Braun meint, dass solche Situationen nur schwer zu durchbrechen und damit geradezu unbeherrschbar seien. Ein Beispiel aus der Schule wäre ein Lehrkollegium, in denen Teilgruppen schon seit Jahrzehnten zerstritten sind und Interventionen unter den Akteuren wenig Aussicht auf Besserung versprechen. Auch hier sind Veränderungen noch möglich, ob diese gelingen, ohne dass z.B. Personen das Team verlassen, ist fraglich. Unvorhersehbare und unbeherrschbare Effekte schließlich sind jene, gegen die im Vorfeld keine Intervention Abhilfe hätte schaffen können.

---

<sup>13</sup> Eine andere Unterscheidung nicht intendierter Effekte bietet Portes (2000): Er unterscheidet zwischen intendierten Ergebnissen, unbeobachteten Ergebnissen, eingangs nicht erwünschten und nicht erwünschten Ergebnissen. Seine Unterscheidung lässt sich weitgehend bei Braun (2008) einfügen, ist aber m.E. noch weniger systematisch, weswegen sich hier auf letzteren bezogen wird.

Für die best-case Studie sind diese Erörterungen insofern relevant, weil der best-case ausgewählt wurde, weil bei ihm vorerst keine auffälligen nicht intendierten, negativen Effekte beobachtet wurden. Für die Studien zum Zentralabitur sind diese Erörterungen wichtig, da sich hier die Frage stellt, inwiefern die Struktur im Bildungssystem bereits für nicht intendierte Effekte verantwortlich ist. Im Weiteren werden nun Ursachen für nicht intendierte Effekte erläutert.

### **3 Kopplungsmechanismen und nicht intendierte Effekte**

Um die Ursachen nicht intendierter Effekte zu systematisieren, werden als Heuristik die beiden in der Prinzipal-Agent Theorie genannten Grundprobleme „hidden information“ und „hidden action“ genutzt (Strøm, 2000, S. 270f). Grob gesagt liegt ersteres vor, wenn die Person, die beeinflussen möchte, über zu wenig oder falsche Informationen verfügt. „Hidden action“ ist zu beobachten, wenn die zu beeinflussende Person nicht die Aufgaben so durchführt wie zuvor verlangt, definiert oder ausgehandelt wurde. Beide Probleme der Steuerung und Governance können durch Kopplungsmechanismen als Teil der Regelungsstruktur ausgelöst wie auch abgemildert werden (vgl. die nächsten Absätze sowie Kapitel 2).

#### *3.1 Ursachen nicht intendierter Effekte*

Für die Beschreibung der Ursachen nicht intendierter Effekte werden als Heuristik die Begriffe „hidden information“ und „hidden action“ aus der Prinzipal-Agent-Theorie genutzt. Sie werden aber zur Anwendung in der Governance-Analyse weiter gefasst. Dies wird nun erläutert.

So kann man mit Braun (2008) die „hidden information“ in zwei Bereiche aufschlüsseln: Das Nicht-Wissen-Können und das Nicht-Wissen-Wollen. Braun (2008) nennt diese Faktoren die kognitiv-psychologischen Ursachen nicht intendierter Effekte. So bezeichnet das „Nicht-Wissen-Können“ (Braun, 2008, S. 107) das unbeabsichtigte Ignorieren von Tatbeständen oder Zusammenhängen. Dieses tritt oftmals bei komplexen Herausforderungen auf, solche finden sich auch im Bildungssystem (vgl. Kapitel 4). Auch das Problem der Informationsasymmetrie, welches die Prinzipal-Agent Theorie zwischen Agent und Prinzipal herausgearbeitet hat, bezeichnet aus Sicht des Prinzipals ein Problem des Nicht-Wissen-Könnens: Denn der Prinzipal weiß nicht im Detail, welche Aufgaben der Agent wie am besten erledigt und zudem ist er nicht in der Lage, das gesamte Verhalten des Agenten zu kontrollieren (vgl. Strøm, 2000, S. 270f). Das Nicht-Wissen-Wollen dagegen bezeichnet strategisches Verhalten, bei welchem bewusst Aspekte des Wissens ausgeblendet werden, da z.B. kurzfristige Interessen langfristige überlagern (vgl. Braun, 2008, S. 108). Dieses könnte im Bildungswesen u.a. in der Bildungspolitik beobachtet werden und kann durch das Problem der „hidden action“ verstärkt werden.

„Hidden action“ bezeichnet den Sachverhalt, dass Steuerung und Governance sich besonders dann schwierig gestalten, wenn die Handlungsziele der zu beeinflussenden Person (Agent) nicht mit den Zielen der Person übereinstimmen, die diesen Agenten

beeinflussen möchte (Prinzipal). Dieser „Extremfall“ wird in der Prinzipal-Agent-Theorie ausführlich debattiert, wobei hierarchische Delegationsbeziehungen im Vordergrund stehen.

So geht die Prinzipal-Agent-Literatur davon aus, dass grundsätzlich zwischen dem Steuerungssubjekt, dem Prinzipal, und dem Steuerungsobjekt, dem Agenten, ein Interessenskonflikt besteht (vgl. z.B. Gilardi & Braun, 2002). Dieses Modell wird zumeist in Bezug auf die Beziehung zwischen Politik und Verwaltung angewendet: Hier wird als Interesse der PolitikerInnen, Erfolg bei einer Wiederwahl genannt („vote seeking“), das Interesse der Verwaltung dagegen sei, eigene Kompetenzen auszubauen („office seeking“) (vgl. Döhler, 2005, S. 217). Bei der Formulierung eines Gesetzes würde daher zwangsläufig die Verwaltung versuchen, vom Gesetz abzuweichen („shirking“). „Das Ausweichverhalten des Agenten und die daraus resultierenden Kontrollprobleme und Interessenkonflikte haben zwangsläufigen Charakter, da grundsätzlich divergierende Präferenzen zwischen Politik und Verwaltung unterstellt werden“, fasst Döhler (2005, S. 216) zusammen. Mittels dieser Annahme der Prinzipal-Agent Theorie wird für die Beziehung zwischen Ministerium und Lehrpersonen ein hohes Konfliktpotential vorausgesetzt, denn Lehrpersonen wird generell ein hoher Autonomiegrad (z.B. im Unterricht) zugestanden (lose gekoppelt, Weick, 1976/ 2009).

Die Stewardship-Theorie bezweifelt diesen Grundkonflikt zwischen Prinzipal und Agent (Donaldson, 1990). Sie geht davon aus, dass überwiegend intrinsische Motivationsfaktoren bei den Agenten vorliegen und die intrinsische Motivation durch z.B. Vertrauen und Teamorientierung aufrechterhalten und nicht durch extrinsische, kontrollorientierte Maßnahmen untergraben werden soll. Laut der Stewardship-Theorie wird Macht durch Expertise ausgeführt und nicht durch normative Vorgaben (vgl. für eine Zusammenfassung: Velte, 2010). Für das Beispiel der Lehrpersonen würde nach der Stewardship-Theorie geschlossen, dass die lose Kopplung durch den gegebenen Vertrauensvorschuss die intrinsische Motivation der Lehrpersonen stärkt und die Beziehung zwischen Ministerium und Lehrperson als Beziehung unter Professionellen gestaltet werden sollte.

Döhler (2008) und Manow (2008) schauen von einem Mittelweg zwischen diesen beiden Theorien auf die Steuerung. So schreiben sie, dass der von der Prinzipal-Agent-Theorie benannte Interessenskonflikt zwar existiert, aber dass er nicht „zwangsläufig“ erfolgt. Dies bedeutet für das Lehrperson-Beispiel: Es besteht die Möglichkeit, dass eine lose Kopplung die intrinsische Motivation von Lehrpersonen stärkt, es besteht aber auch die Möglichkeit, dass der gegebene Freiraum von einigen Lehrpersonen dazu genutzt wird, von den Leistungsvorgaben abzuweichen (aus welchen Gründen auch immer). Kopplungsmechanismen sollten demnach aus Sicht des Prinzipals beide auftretenden Möglichkeiten abdecken, d.h. auf beide Situationen reagieren können. U.a. Manow (2008) und Döhler (2008) zeigen auch anhand von empirischen Studien hinsichtlich unterschiedlichster Akteure auf, wie Kopplungsmechanismen gestaltet werden können, die den Interessenskonflikt zwischen Agent und Prinzipal abmildern können (vgl. Kapitel 3.2.2).

Schließlich hebt Braun hervor (2008, S. 121), dass „hidden action“ bereits durch das Handlungssystem, in dem Akteure agieren, entstehen kann: „Je mehr Akteure mit-

spielen, auf je mehr Ebenen die Handlungen stattfinden und je interdependenter das Handeln ist, umso schwieriger wird es für die Akteure – ganz unabhängig von den individuellen Dispositionen –, die Zukunft zu berücksichtigen.“ Das bedeutet im Umkehrschluss, dass nicht intendierte Effekte durch Rahmenbedingungen trotz „hidden action“ und „hidden information“ eingegrenzt werden können, sie können indes nicht abschließend getilgt werden.

### 3.2 Reduzierung nicht intendierter und Forcierung intendierter Effekte

Im Folgenden wird dargelegt, wie mit den oben beschriebenen nicht intendierten Effekten, die in den Studien zum Zentralabitur untersucht wurden, umgegangen werden kann. Nicht intendierte Effekte zu reduzieren, ist zudem wichtig, um intendierte Effekte hervorzubringen. Denn Nebeneffekte können auch die gesetzten Ziele untergraben oder ins Gegenteil umkehren (vgl. Börs, 2002).

#### 3.2.1 Kopplungsmechanismen als Antwort auf fehlende Information und „hidden information“?

Wie kann mit nicht intendierten, besonders unerwünschten Effekten, umgegangen werden? Grundlegend können nicht intendierte Effekte nicht abgeschafft werden, da die Produktion von Wissen gleichzeitig immer Nichtwissen koproduziert (z.B. Braun, 2008; Wehling, 2008). Durch erhöhte Koordination kann aber versucht werden, nicht intendierten, unerwünschten Effekten entgegenzuwirken. Braun (2008, S.103f) schreibt, dass durch „ein verstärktes Bemühen um einen reflexiven Umgang sowohl mit den möglichen Nebenfolgen politischen Handelns als auch mit möglichen kollektiven Dilemmas“ versucht werden kann, unerwünschte Effekte zu reduzieren (vgl. Braun, 2008, S. 103f). Das heißt, dass hinsichtlich der Kopplungsmechanismen ex ongoing Mechanismen sinnvoll sein können, um nicht intendierte Effekte zu reduzieren. Durch ex ongoing Mechanismen werden Prozesse begleitet. Z.B. können frühzeitig Fehlerquellen durch Monitoring identifiziert und behoben werden. Eine konstruktive Fehlerkultur, aufbauend auf ex ongoing Mechanismen, könnte somit eine Antwort auf unerwünschte Effekte im Bildungssystem sein.

Ein System, welches Unterstützung bei Fehlern anbietet, muss sich auch gegen Akteure wappnen, die von der Norm abweichen („shirking“). Denn es wird mit Döhler (2008) und Manow (2008) angenommen, dass abweichendes Verhalten zwar nicht die Regel ist, aber in der Praxis auftritt. Osborne (2004) beschreibt im Rahmen der Spieltheorie mit der Argumentationsfigur der „glaubwürdigen Drohung“, dass Sanktionen, also ex post Mechanismen, auch ohne im Normalfall angewendet zu werden, antizipatives Verhalten auslösen können.<sup>14</sup> Voraussetzung dafür ist, dass im tatsächlichen Fall einer Normverletzung die Sanktion angewendet wird, denn ansonsten handelt es sich

---

<sup>14</sup> Hier zeigen sich Parallelen und Differenzen zwischen Spieltheorien und Lerntheorien. So ist die Argumentationsfigur der „glaubwürdigen Drohung“ (vgl. Osborn, 2004) mit Aspekten der behavioristischen Lerntheorie vergleichbar; das Konzept der „Fairness“ in der neueren Forschung zur Spieltheorie ist jedoch eher mit u.a. der kognitivistischen Lerntheorie zu verbinden (vgl. u.a. Sauer mann & Kaiser, 2010).

nicht um eine „glaubwürdige“ Drohung, sondern lediglich um einen „Papiertiger“. Das heisst, um eine Verordnung ohne Auswirkungen.

### 3.2.2 Kopplungsmechanismen als Antwort auf „hidden action“?

Im Folgenden wird dargelegt, inwiefern ex ante, ex ongoing und ex post Kopplungsmechanismen das Auftreten von „hidden action“ reduzieren können und die Chance erhöhen, dass die Agenten die Ziele ihrer Prinzipale umsetzen. Aus Platzgründen können an dieser Stelle nicht alle Kopplungsmechanismen und auch nicht ihre Kombinationen thematisiert werden (vgl. Tabelle 1). Es wird sich daher auf die m.E. bisher am besten erforschten Kopplungsmechanismen fokussiert; ex ante, ex ongoing und ex post Kopplungsmechanismen. Im Ansatz wird reflektiert, ob sie als eher formell, informell, strukturell, relational oder weltanschaulich eingeordnet werden können.

#### *Ex ante und ex ongoing Mechanismen*

Um „hidden action“, also „verborgene Handlungen“ die durch Interessenskonflikte zwischen Agent und Prinzipal hervorgerufen werden, zu verhindern, bieten sich ex ante Mechanismen an. Sie umfassen z.B. die sorgfältige Auswahl der Agenten. Ex ante Mechanismen sollen schon vor der Leistungserbringung sicher stellen, dass kein bzw. ein geringer Interessenskonflikt zwischen Prinzipal und Agent vorliegt. Ex ongoing Mechanismen bauen ebenfalls auf der Idee auf, Interessenkonflikte zwischen Prinzipal und Agent durch soziale Einbindung und Konsensherstellung zu reduzieren (Döhler, 2005; Manow, 2005). Das theoretische Argument ist, dass die soziale Einbindung und der Konsens Vertrauen schafft, welches sich positiv auf die Gestaltung der Beziehung auswirkt. Als positive Beispiele für Kopplungsmechanismen, die dem Agenten Vertrauen gegenüber aussprechen und somit Konflikte reduzieren können, nennt Döhler (2005, S. 217 und S. 225) u.a. in der Verwaltung das „Autopilotprogramm“, welches immer dann greift, wenn Veränderungen stattfinden, wenn z.B. ein Vorgesetzter pensioniert wird. Im Autopilotprogramm darf die Verwaltung die Routine selbst herstellen. Dies funktioniert, wenn das „administrative Mandat“, welches auf dem Professionsverständnis der Verwaltung als neutrale Instanz basiert, als gültig anerkannt wird (vgl. Döhler, 2005, S. 221-224).

Weitere Beispiele aus der Politikwissenschaft, die zur Einbettung der Akteure beitragen und somit Konflikte zwischen Prinzipal und Agent reduzieren sollen, sind u.a. der Mechanismus der „compliance“ (Döhler, 2005, S. 219). Dieser bedeutet, dass die Agenten mit ihrer Expertise sich an der Formulierung von Richtlinien beteiligen. Durch die Mitwirkungsmöglichkeiten werden die Interessen des Prinzipals zu den Interessen der Verwaltung und Reibungsflächen können verringert werden. Auch Karriereverflechtungen sind ein Mittel um Personen an die Organisation zu binden (Manow, 2005, S. 256).

Einen weiteren Aspekt, der in der Studie A (best-case Analyse) relevant wird, stellt die Netzwerktheorie heraus. Sie hält fest, dass multiplexe Kopplungen, also mehrsträngige Kopplungen unterschiedlicher Art (formal und relational etc.) besonders gut geeignet sind, dass ein Individuum im Bedarfsfall Unterstützung erhält (Jansen, 2006, S. 108). Analog kann festgehalten werden, dass multiplexe Kopplungen geeignet

sind mit einer hohen Wahrscheinlichkeit eine bestimmte Information weiterzuleiten. Es ist z.B. sicherer, dass eine Information ankommt, wenn sie im Professionsverständnis enthalten ist, durch Vorgesetzte ausgesprochen wird und von FreundInnen anerkannt wird. Andererseits kann solch eine multiplexe Bindung so eng sein, dass sie Innovationen verhindert (vgl. Jansen, 2006, S. 31f).

Die Theorie zu Netzwerkanalysen hebt auch die Schwächen von sozialer Einbindung hervor, so steigt in dichten Netzwerken der Druck zur Konformität, was sich je nach Zielsetzung der Gruppe negativ wie positiv auswirkt. Ein bekanntes Beispiel ist die Mafia. Sie wählt ihre Mitglieder genau aus, bindet sie stark sozial ein und verfolgt selbstgesetzte Ziele, diese sind jedoch rechtstaatlich und politisch nicht erwünscht. Dichte Netzwerke, die durch ex ante und ex ongoing Mechanismen gefördert werden, können sich demnach auch negativ auswirken. Signale für solch negative Entwicklungen sind nach Meyer und Rowan (1977, S. 359) in Organisationen wie dem Bildungssystem, wenn Einblicke von außen unerwünscht sind. Diese können nach Meyer und Rowan (1977) Evaluationen und Inspektionen beinhalten, die die „Logik des Vertrauens“ stören könnten. Diese Abschottung kann, so die Autoren, Innovationen verhindern und Fehlverhalten verschleiern. Andererseits erscheint Vertrauen nach Döhler (2005) eine Voraussetzung, um Konflikte und abweichendes Verhalten zu verhindern. Ein Mittelweg zwischen Vertrauen und Kontrolle, scheint hier wünschenswert.

#### *Ex post Mechanismen*

Ex post Mechanismen entsprechen extrinsischen, zumeist monetären Anreizen, wie Boni oder Auszeichnungen für eine erfolgreiche Umsetzung oder aber auch Sanktionen bei einer Nicht-Erreichung von Zielen. Sie zielen nicht auf argumentative Überzeugung oder soziale Einbindung wie die ex ongoing oder ex ante Mechanismen, sondern bieten extrinsische Anreize. Die Idee ist, dass u.a. Boni vom rational abwägenden Agenten in seine/ihre Kosten/Nutzen-Kalkulation einbezogen werden. So soll eine Übereinstimmung zwischen Ziel und Umsetzung sichergestellt werden. Ex post Mechanismen bieten den Agenten Gestaltungsspielraum für die Anpassung von Mitteln zur Zielerreichung (Schedler & Proeller, 2006).

Hinsichtlich der Wirkungsweise von ex post Kopplungsmechanismen im Bildungssystem, wie es in formaler Hinsicht das Zentralabitur in Hessen und Bremen darstellt, stellen Frey und Jegen (2002, S. 30) heraus, dass ex post Mechanismen bei der Bereitstellung öffentlicher Güter, die Gefahr der „Übeneralisierung des pay-for-performance-Prinzips“ bergen (vgl. auch Studie C der Dissertation). Übeneralisierung heißt, dass der Agent nur die vom Prinzipal definierten Aufgaben erfüllt, für die er ex post honoriert wird bzw. für die ihm Sanktionen bei Nicht-Erfüllung der Aufgabe drohen.

„Aus diesem Grund führen extensive Kontrollen oder Anreizkontrakte häufig zu unerwünschtem und für die Firma schädlichem Verhalten. Die Mitarbeitenden konzentrieren sich ausschließlich auf die monetär belohnten Tätigkeiten, vernachlässigen hingegen diejenigen, für die sie kein monetäres Entgelt erhalten“ (Frey & Jegen, 2002, S. 35).

So weisen Frey und Jegen (2002, S. 32) darauf hin, dass ex post Mechanismen, die die extrinsische Motivation stärken, intrinsische Motivation verdrängen können (vgl. „Mo-

tivation Crowding Theory“). Bedingung für das Auftreten dieser Übergeneralisierung bzw. dieses Verdrängungseffektes ist u.a. dass der ex post Mechanismus die Selbstbestimmung der Mitarbeitenden einschränkt (Frey & Jegen, 2002, S. 32). Unterstützende Mechanismen, also ongoing Kopplungsmechanismen, können diesen Effekt wieder kompensieren (vgl. Frey & Jegen, 2002, S. 32).

#### *Zusammenfassung*

Hinsichtlich des Bildungssystems weist die Literatur theoretische Argumente für einen Einsatz von ex ante und ex ongoing Mechanismen auf. In Bezug auf ex ongoing Mechanismen wie z.B. Monitoring, Feedback, etc. werden in der Literatur positive Ergebnisse erwartet (u.a. Bos & Schwippert, 2002). Ex post Mechanismen dagegen werden in Bezug auf das Bildungssystem eher kritisch gesehen, da diese die intrinsische Motivation der Mitarbeitenden verdrängen können (vgl. Frey & Jegen, 2002).

Ex ante wie ex ongoing Kopplungsmechanismen sind aber nicht nur positiv: So können ex ante Mechanismen zwar „hidden action“ durch Einbindung der Akteure abschwächen, sie machen aber nicht sichtbar, ob und in welcher Weise die erforderte Leistung von den Akteuren tatsächlich erbracht wird (vgl. u.a. Meyer & Rowan, 1977). Die Ergebnisse aus PISA zeigten, dass die Input-Steuerung nur begrenzt erwünschte Erfolge erzielt (OECD, 2007, 2010). Ex ongoing Mechanismen dagegen bergen die Gefahr, dass der Druck zur Konformität so stark wird, dass Innovationen verhindert werden (vgl. Netzwerktheorie, u.a. Jansen, 2006).

Nachdem nun Kopplungsmechanismen im Zusammenhang mit (nicht) intendierten Effekten erörtert und als Möglichkeit diskutiert wurden, „hidden information“ wie „hidden action“ abzuschwächen, wird im Folgenden das Bildungssystem mit seiner Eigenlogik vorgestellt. Dies dient dazu, die Kopplungsmechanismen kontextspezifisch für das Bildungssystem analysieren zu können und die Fragestellungen der Synopse herzuleiten (Lasky, Datnow & Stringfield, 2005).

## **4 Was ist ein Bildungssystem?**

Im Folgenden wird das Bildungssystem organisatorisch, hinsichtlich seiner Aufgaben und Ziele (Funktionen) sowie in Bezug auf seine Besonderheiten dargestellt. Eigenlogiken des Systems werden identifiziert, um die mögliche Wirkung von Kopplungsmechanismen im Bildungssystem besser einordnen zu können.

### *4.1 Die Schule als Organisation*

Schule wird oft von ihren Mitgliedern als organisationshaft erlebt (Drepper & Tacke, 2012, S. 206). Dies liegt vor allem daran, dass Schule folgende organisationstypische Charakteristika aufweist (Drepper & Tacke, 2012, S. 206):

- Kollektivcharakter und Überindividualität (Sozialdimension: Kinder werden zu SchülerInnen in Gruppen zusammengefasst. Dies produziert kollektives Rollenleben)



- Regelmäßigkeit und Formalität (Sachdimension: Richtlinien, Vorschriften, Gesetze und Konventionen gelten, Hierarchien bestehen (Themen werden vorgegeben), Mitgliedschaften wie Ein- und Austritt sind definiert)
- Regelmäßigkeit und Wiederholung (Zeitdimension: Zeiteinteilung in Stundenpläne, Jahrgangsstufen, Schuljahre,...)
- Orts- und Raumbezug (Raumdimension: spezifische Räume werden über lange Zeiträume bereitgehalten und man kann dort bestimmte Personen erwarten)

Organisatorisch bestehen Schulen einerseits aus der inneren Organisation (u.a. Unterricht, Lehrpersonenkollegium, Schulleitung, Sekretariat, HausmeisterIn) und andererseits aus einer äusseren Organisation, dem bürokratischen Schulsystem (Drepper/ Tacke 2012: 209). Innerhalb der inneren wie äusseren Organisation befinden sich unterschiedliche Delegationsketten: So gibt es innerhalb der Schule eine drei- bis vierstufige Hierarchie von der Schulleitung, (Stellvertretung), der Lehrperson und der/dem SchülerIn (vgl. Klafki, 2002, S. 125). Hinzu kommen Einflussmöglichkeiten der Eltern oder SchülerInnen durch Partizipationsmöglichkeiten (z.B. Elternrat, SchülerInnenvertretung) (vgl. Klafki, 2002, S. 125).

Für die äußere Organisation hält Strøm (2000, S. 267) aus politikwissenschaftlicher Perspektive für repräsentative, zentral organisierte Demokratien fest, dass es sich um vierstufige Delegationsketten und demnach um vier zu überbrückende Schnittstellen handelt: (1) Von der WählerIn zu den gewählten RepräsentantInnen, (2) Von der Legislative zur Spitze der Regierung, (3) Von dort zu den verschiedenen Vorgesetzten der Exekutive und (4) von den Vorgesetzten der Exekutive zu den Beamten. Da sich die Dissertation mit Deutschland und Kanada, zwei föderal organisierten Staaten, beschäftigt, muss diesen Delegationsketten zum Einen noch der Bund als Kooperationspartner (und damit teilweise als zweiter Delegationsstrang) hinzugefügt werden und zum zweiten endet die Delegationskette nicht bei den Beamten, sondern sie geht weiter bis zur Schulleitung und der inneren Organisation der Schule. D.h. die von Klafki (2002) erwähnten drei bis vier Stufen der Autoritätshierarchie müssen ebenfalls noch ergänzt werden, daraus folgen rund sieben bis acht Schnittstellen: (5) von den Beamten zu der Schulleitung, (6) von der Schulleitung (bzw. der Stellvertretung) zu den Lehrpersonen sowie (7) von den Lehrpersonen über den zu unterrichtenden Stoff zu den SchülerInnen (vgl. auch Gilardi & Braun, 2002; Reusser, 2009).

Dies beschreibt die für das Bildungswesen charakteristische Mehrebenenstruktur (vgl. auch Altrichter & Maag Merki, 2010). In dieser agiert die Einzelschule als „pädagogische Handlungseinheit“ (Fend, 1988). Rolff (2007, S. 44) schreibt: „Das Hauptproblem des Gesamtsystems besteht in der Kopplung der Entwicklung der Einzelschulen (‚Selbststeuerung‘) mit der Entwicklung des Gesamtsystems (‚Systemsteuerung‘).“

Ein wichtiges Merkmal, welches sich daraus für Schulen ergibt, ist, dass Schulen ihre Identität nicht selbst definieren, sondern dass ein wesentlicher Teil „ihrer jeweiligen Identität im Spannungsfeld und in der Auseinandersetzung mit ihrer staatlichen und professionellen Rahmenordnung“ festgelegt wird (Drepper & Tacke, 2012, S. 231). Darüber hinaus, sind besonders auch gesellschaftliche Aushandlungsprozesse für die

Definition der Aufgaben von Schulen und Bildungssystem relevant (vgl. nächster Abschnitt).

#### *4.2 Funktionen von Schule und Bildungssystem*

Im Rahmen der Theorie der Schule werden diese Fremd- und Selbstzuschreibungen diskutiert. Blömeke, Herzig und Tulodziecki (2007, S. 74) resümieren in ihrem Überblick zur Theorie der Schule, dass „eine umfassende Theorie der Schule, die unter Berücksichtigung aller bedeutsamen Perspektiven die gesellschaftlichen Funktionen und die Binnenstruktur dieser Institution beschreibt und erklärt, derzeit nicht existiert.“ Sie verorten die bisher existierenden Theorien auf der Makro- und Mesoebene des Bildungssystems und problematisieren, dass bisher keine Theorie existiert, die beide Ebenen aufeinander bezieht. Denn nur so könne ein umfassendes, differenziertes Bild von Schule entwickelt werden (Blömeke, Herzig & Tulodziecki, 2007, S. 55f).

Theorien, die die Makroebene berücksichtigen, stellen Blömeke, Herzig und Tulodziecki (2007, S. 59-67) zufolge das Verhältnis Schule und Gesellschaft in den Mittelpunkt. Auf der Mesoebene werden dagegen die internen Strukturen der Schule analysiert. Als Beispiele führen die AutorInnen psychologische und interaktionistische Theorien an (vgl. Blömeke, Herzig & Tulodziecki, 2007, S. 67-73) sowie die Schuleffektivitätsforschung (vgl. Blömeke & Herzig, 2009, S. 19).

Da eine Zusammenschau der bisherigen Theorien der Schule, inhaltlich hier nicht diskutiert werden kann, werden an dieser Stelle kurz Fends Theorie der Schule und die Schuleffektivitätsforschung als Beispiele für viel diskutierte Theorien der Makro- und Mesoebene angeführt (vgl. für eine Übersicht u.a. Blömeke & Herzig, 2009).

Fend (2006) neue Theorie der Schule unterscheidet zwischen den gesellschaftlichen Funktionen und den individuellen Funktionen des Bildungssystems (vgl. Fend, 2006, S. 54). Zu den gesellschaftlichen Funktionen gehören die Enkulturation, die Qualifikation, die Allokation und die Legitimation und Integration. Die individuellen Funktionen setzen sich zusammen aus der kulturellen Teilhabe und Identität, der Berufsfähigkeit, der Lebensplanung und der sozialen Identität und politischen Teilhabe. Insgesamt verfolgen Gesellschaften mit der Einrichtung von Bildungssystemen in der Moderne nach Fend (vgl. Fend, 2006, S. 49-55) gleichzeitig die Ziele soziale Reproduktion und Innovation. Unter Berücksichtigung des Faktums, dass die Funktionen von Schule in Abhängigkeit zur Gesellschaft definiert werden, verwundert es nicht, dass die genannten Funktionen Ergänzungen und unterschiedliche Konkretisierungen erfahren (vgl. Beispiele bei Blömeke, Herzig & Tulodziecki, 2007, S. 63) Berkemeyer und Manitus (2013, S. 232f) arbeiten z.B. eine gerechtigkeits-theoretische Perspektive in ihre (auf Fend bezugnehmenden) schultheoretischen Überlegungen ein.

Obwohl die Schuleffektivitätsforschung sich nicht explizit auf Theorien der Schule bezieht, nehmen Blömeke und Herzig (2009, S. 19f) sie mit in den Kanon der Schultheorien auf. Denn auf der Basis von SchülerInnenleistungen werden induktiv Funktionen von Schule, nämlich „die Effektivität der Institution Schule empirisch zu modellieren versucht“. Auch dies kann als eine Erweiterung von Funktionen von Schule angesehen werden.

Blömeke, Herzig und Tulodziecki (2007, S. 63) referieren zudem, dass die Funktionsbeschreibungen von Schule in ihren Konkretisierungen umstritten sind (z.B. in welcher Form ist Selektion erwünscht?), zudem sind den genannten Funktionen von Schulen Widersprüche immanent. So ist z.B. die Frage, wie Integration und Selektion konkret gemeinsam zu erreichen sind. Blömeke, Herzig und Tulodziecki (2007, S.74) schließen, dass die Funktionen von Schule als Hintergrundfolie für Analysen dienen können, vor der schulische und unterrichtliche Fragen eingeordnet und Spannungsverhältnisse und deren Auflösung differenziert analysiert werden können.

#### 4.3 Spezifika des Bildungssystems

In der erziehungswissenschaftlichen Literatur wird besonders die Technologievagheit (Bellmann & Weiss, 2009, S. 297), die damit zusammenhängende lose Kopplung der Lehrpersonen (Weick, 1976/ 2009), und die Rekontextualisierung von Vorgaben (Fend, 2008) betont. Das folgende Zitat von Bellmann und Weiss (2009, S. 297) verdeutlicht diese drei Spezifika und ihren Zusammenhang:

„Der Lehrerberuf war bislang durch hochgradig unvollständige Verträge und einen entsprechend hohen Vertrauensvorschuss gekennzeichnet. Dies gilt auch für andere Berufe, bei denen der kreative Umgang mit Wissen und Können im Vordergrund steht und die durch komplexe Zielstellungen und vage Technologien gekennzeichnet sind.“

Im Lehrerberuf steht die Interaktion zwischen Lehrenden und Lernenden im Mittelpunkt. Diese kann jedoch nicht vollständig im Vorhinein durch z.B. Verträge festgelegt werden, da situativ wie individuell reagiert und agiert werden muss. Dies impliziert eine Technologievagheit, die in anderen Berufen (z.B. Maschinenbau) in diesem Maße nicht vorliegt. Die Lehrpersonen (u.a.) sind daher nur lose gekoppelt mit der Schulleitung bzw. dem Ministerium, da im Unterricht die Lehrperson in besonderer Weise ihre eigene Identität einbringt (vgl. Weick, 1976/ 2009, S. 88). Weick (1976/ 2009, S. 92) definiert als Grad der Kopplung die der Handlung vorausgehenden Voraussetzungen und schreibt: „je länger die Kette von Voraussetzungen, desto enger die Kopplung“. Dies bedeutet gleichzeitig, dass Vorgaben übersetzt und angepasst werden müssen. Fend (2008, S.174f) fasst diese Übersetzungsleistung als Rekontextualisierung. Dies besagt u.a., dass Auftragshandeln in der Schule individuell ausgestaltet und adaptiert werden muss (vgl. Fend, 2008, S. 175).

Eine Beschreibung des Bildungssystems kommt ohne diese Spezifika nicht aus, doch werden die Befunde in der Erziehungswissenschaft m.E. bisher nicht unter einem Begriff subsummiert. Die Recherche zur Organisation Schule zeigte nun eine interessante inhaltliche Übereinstimmung zwischen organisationssoziologischen und erziehungswissenschaftlichen Ergebnissen zur Theorie der Schule, weswegen im Folgenden der Begriff der „sozialen personenbezogenen Dienstleistungsorganisation“ (Drepper, 2010, S. 143) (SPDO) eingeführt wird. Drepper und Tacke (2012, S. 210):

„Was erkennbar wird, ist eine Wahlverwandtschaft bzw. Analogie zwischen den organisationstheoretischen Einsichten der letzten Jahrzehnte über schulische Organisationen und den pädagogischen Befunden zu unterrichtenden und unterrichteten Menschen: Organisationstheoretiker haben für Organisationen herausgefunden, was Pädagogen in Bezug auf Lernsubjekte immer schon gewusst haben. Von da aus liegt dann der schultheoretische Schluss nicht fern, die Schule als eine `pädagogische Handlungseinheit` (Fend 1986) zu be-

zeichnen, denn gerade Erziehungsorganisationen zeigen die organisationstypischen Charakteristika der Nicht-Trivialität, der externen Nicht-Steuerbarkeit, Eigensinnigkeit und Eigenlogik im besonderen Maße.“

Die Ergebnisse der Erziehungswissenschaft zum Bildungssystem zeigen Eigenlogiken auf (Technologievagheit, lose Kopplung und Rekontextualisierung), die auch in der Definition des Bildungssystems als soziale personenbezogene Dienstleistungsorganisation (Eigenlogik, Nicht-Steuerbarkeit und Eigensinnigkeit) sichtbar werden. Diese Charakteristika des Bildungssystems stellen spezielle Anforderungen an Governance im Bildungsbereich, diese werden anhand des Konzeptes der SPDO im Folgenden näher erläutert.

Das Bildungssystem bezeichnet Drepper (2010, S. 143) als eine soziale personenbezogene Dienstleistungsorganisation (SPDO), da ihr „dominanter Aufgabenbezug das Individuum in unterschiedlichen Situationen und Problemlagen“ ist und die folgenden sechs Struktureigenschaften von SPDOs erfüllt (vgl. Hasenfeld, 1983, zitiert nach Drepper, 2010, S. 144):

1. Personen sind das „Rohmaterial“ (Input) und veränderte Personen bzw. veränderte Verhaltensweisen von Personen das Output
2. „Ziele und Zwecke sind oftmals mehrdeutig und interpretativ umkämpft“
3. SPDOs „agieren in vielschichtigen und komplexen Umwelten, in denen verschiedene Anspruchsgruppen ihre Interessen und Wertansprüche an die Organisationen richten“
4. Kerntätigkeiten laufen über interaktive Mitarbeiter-Klienten-Beziehungen
5. Die „Interaktionsarbeit bedeutet hohe Komplexität und unsichere Ergebnisse bzw. Fallausgänge, da Arbeit an und mit Personen als nicht-trivialer Prozess zu verstehen ist und damit einen hohen Unsicherheitsaspekt aufweist“ (Technologiedefizit (vgl. Luhmann & Schorr, 1979, zitiert nach Deppert, 2010, S. 144))
6. Durch das Technologiedefizit entsteht zusätzlich auch ein Effizienz- und Effektivitätsdefizit (vgl. für alle sechs Charakteristika: Hasenfeld, 1983, zitiert nach Drepper, 2010, S. 144)

Im Bildungssystem sind sozusagen die SchülerInnen das Rohmaterial, die Ziele der Bildung sind komplex, da sie von Bildungsinhalten über Sozialverhalten bis hin zu demokratischer Erziehung reichen und sie sind zudem umkämpft (vgl. Blömeke, Herzig & Tulodziecki, 2007, S. 63). Dies zeigt sich z.B. an der Diskussion um die Abschaffung des Gymnasiums oder der Einführung von Gesamtschulen. Hier wird nicht alleine die konkrete Umsetzung oder Effizienz eines Instrumentes verhandelt, sondern es werden Wertvorstellungen definiert und diskutiert. Das bedeutet, dass Steuerung in und von sozialen personenbezogenen Dienstleistungsorganisationen ungleich schwieriger ist als die Steuerung innerhalb eines Unternehmens, da ein Wertekonsens in der Gesellschaft gefunden werden muss, um die SPDOs und damit auch die Arbeit in den Schulen zu legitimieren. Die Kernarbeit an einer Schule verläuft zudem zwischen Lehrperson und SchülerIn, aber auch zwischen Lehrpersonen und Lehrpersonen oder Lehrperson und Schulleitung (vgl. interaktionistische Theorien zur Schule, u.a. Mead, 1969 und Ulich, 2001; verwiesen nach Blömeke, Herzig & Tulodziecki, 2007, S. 68f). Interaktionen sind somit Voraussetzung für die Zielerreichung (Technologievagheit, vgl. Bellmann &

Weiss, 2009, S. 297). Schulen und das Bildungssystem sind in dessen Folge auch mit dem Vorwurf der Ineffizienz und Ineffektivität konfrontiert (vgl. Schuleffektivitätsforschung, vgl. u.a. Bos & Schwippert, 2002).

Diese Besonderheiten können nicht durch Strukturveränderungen gelöst werden, sie sind strukturimmanent: „Für Schulen ist damit scheinbar die Gleichzeitigkeit von formalen Strukturen und schwer – bis nicht formalisierbarer Situativität bzw. Ereignishaftigkeit [u.a. im Unterricht, Anmerkung von DJJ] kennzeichnend“ (Drepper & Tacke, 2012, S. 210). Die verstärkte Unsicherheit und Unplanbarkeit fordert vom Bildungssystem, wie auch z.B. im Politikfeld Gesundheit andere Strategien der Governance und Kopplungsmechanismen als Situationen in Unternehmen wie z.B. der Automobilindustrie (vgl. u.a. Langer, 2009b, S. 153).

#### *4.4 Implikationen für die Governance – Empirische Befunde*

In der Empirie zeigen sich in sozialen personenbezogenen Dienstleistungsorganisationen (wie im Bildungs- und Gesundheitssystem) zwei sich ergänzende Strategien (vgl. Drepper, 2010): Zum einen sind die Ebenen und Schnittstellen lediglich lose gekoppelt (Weick, 1976/ 2009), so können die Akteure – wie z.B. die Lehrpersonen im Unterricht – auf die Herausforderung der Situativität reagieren. Zum anderen wird in der Literatur die Annahme aufgestellt, dass das zentrale Merkmal von Bildungssystemen nicht Effizienz und Effektivität, sondern die Fähigkeit ist, gesellschaftliche Werte zu legitimieren (Drepper, 2010, S. 147). Dies bedeutet, dass es die Hauptaufgabe für Erziehungsorganisationen ist „Vertrauen und Zutrauen der Öffentlichkeit in die Sinn- und Werthaftigkeit der eigenen Aufgabe und Tätigkeiten zu gewährleisten und auf Dauer zu stellen“ (Drepper, 2010, S. 147). Das heißt, dass die komplexen Aufgaben, wie sie typisch für soziale personenbezogenen Dienstleistungsorganisationen sind, nicht durch Verträge umfassend definiert werden können. Sie können damit nur delegiert und bewältigt werden, wenn sie von den Akteuren „übersetzt“ werden. Die Lehrperson muss im Unterricht die Vorgaben des Ministeriums auslegen. Übersetzung bedeutet also zugleich auch Interpretation der Vorgaben. Voraussetzungen für die Interpretation der vorgegebenen Ziele sind Gestaltungsspielräume der z.B. Lehrpersonen und seine/ihre intrinsische Motivation, diese Übersetzungsleistung auch zu übernehmen. Lose Kopplungen, also *formal* lose Kopplungen, da sie nicht an Hierarchien gebunden sind, können diesen Spielraum ermöglichen.

Um die Interpretationen jedoch zu kanalisieren, verpflichtet die Instanz, die die Ziele vorgibt, die Angestellten der SPDOs, die Wertigkeit der eigenen Aufgaben gegenüber der Gesellschaft offen zu legen. Dies soll dazu führen, dass die komplexen Aufgaben der SPDOs nicht individuell interpretiert werden, sondern professionell im Sinne der zielvorgebenden Instanz. Dies entspricht einer weltanschaulichen Kopplung (Lasky, Datnow & Stringfield, 2005), die *ex ante* wie *ex ongoing*, also vor und während der Leistungserbringung, erzeugt wird.

Die beiden Strategien (lose und weltanschauliche Kopplung) ergänzen sich, da durch die Werterzeugung sichergestellt wird, dass das lose gekoppelte System nicht „auseinanderfällt“.

In den letzten Jahren aber wurden im Bildungssystem – wie in der Einleitung erwähnt – auch die Kopplungsmechanismen reformiert. So wurden ex post Instrumente wie das Zentralabitur und ex ongoing Instrumente wie Schulinspektionen eingeführt. Diese Entwicklungen führen zu den Fragestellungen der Synopse.

## **5 Zwischenfazit und übergeordnete Fragestellungen zu den Studien**

Die Arbeit stellt bisher heraus, dass Kopplungsmechanismen Teil der Regelungsstrukturen von Governance und Steuerung sind und dass von ihnen erwartet wird, dass durch sie Leistungen u.a. in Bildungssystemen variieren können. So wurde hervorgehoben, dass Kopplungsmechanismen dazu beitragen, (nicht) intendierte Effekte auszulösen. Zudem wurde referiert, dass in Bildungssystemen bis Anfang des Jahrtausends hauptsächlich versucht wurde durch weltanschauliche Kopplung (Lasky, Datnow & Stringfield, 2005), die ex ante wie ex ongoing, also vor und während der Leistungserbringung erfolgte, sicherzustellen, dass die Akteure die vorgegebenen Ziele des Bildungssystems erfüllen. Das bedeutet, in Bildungssystemen wurde durch ex ante Mechanismen (wie der Auswahl der Lehrpersonen) versucht, zu garantieren, dass die Akteure das tun, wozu sie angestellt sind (vgl. Meyer & Rowan, 1977, S. 360). Seit der Veröffentlichung von Daten zur Performanz von Bildungssystemen sind Bildungssysteme und ihre Akteure aufgrund der dort vorgefundenen Mängel (vgl. Bos & Schwippert, 2002) nun konfrontiert mit ex post Mechanismen, wie den zentralen Abschlussprüfungen, sowie mit (teilweise extern durchgeführten) Feedbackschlaufen und Monitoring (ex ongoing Kopplungsmechanismen). Die Kopplungsmechanismen wurden also reformiert. Daraus ergeben sich für die Forschung folgende Fragen:

- I. Welche Kopplungsmechanismen können dazu beitragen, dass in Bildungssystemen Ziele umgesetzt werden, das heißt, die erwünschte Leistung erbracht wird?
- II. Wie wirken sich Reformen der Kopplungsmechanismen aus?
  - a. Welche positiven Effekte können erzielt werden?
  - b. Welche negativen Effekte ergeben sich?

Zu Frage I wurde mit Döhler (2008) und Manow (2008) argumentiert, dass Kopplungsmechanismen kombiniert eingesetzt werden sollten. So wird angenommen, dass die Mehrheit der Akteure über intrinsische Motivation verfügt und diese nicht durch extrinsische Kontrollmechanismen geschwächt werden sollte (vgl. Frey & Jegen, 2002). Dies spricht für einen Einsatz von ex ante Kopplungsmechanismen. Da aber auch abweichendes Verhalten – bewusst oder unbewusst vollzogen – von Akteuren erwartet wird, sollten Systeme auch Kopplungsmechanismen für den Fall des „shirking“ bereithalten. Hier bieten sich besonders ex ongoing Mechanismen an, da diese mitberücksichtigen, dass gerade in sozialen personenbezogenen Dienstleistungsorganisationen mit nicht intendierten Effekten gerechnet werden muss und daher Methoden des Umgangs mit unerwünschten Effekten bereitgestellt werden sollten, die nicht zu einer Demotivation der Akteure führen. Zudem – so wird argumentiert – sollten Systeme auch gegenüber strategischem Verhalten, welches zu nicht intendierten Effekten

führt („Nicht-Wissen-Wollen“, Braun, 2008) gewappnet sein. Hier könnten ex post Mechanismen greifen.

Die beiden übergeordneten Fragen, die theoretisch in den vorliegenden Kapiteln diskutiert wurden, beinhalten zwei Teilfragen, die in den vorliegenden Studien empirisch untersucht werden:

1. Mittels welcher Kopplungsmechanismen wird in einem best-case gesteuert? (Studie A)
2. Erzeugen ex post Kopplungsmechanismen nicht intendierte Effekte? (Studie B, C und D)

Die Arbeit liefert damit Ergebnisse zu Teilaspekten der Frage I und der Frage II b. Im Folgenden werden nun die Erwartungen, Methoden und Ergebnisse der Studien A, B, C und D präsentiert. Die Ergebnisse werden anschließend in der Diskussion geordnet nach der best-case Studie (Studie A) und den Zentralabitur- Studien (Studie B, C und D) sowie gemeinsam diskutiert.

## **6 Erwartete Kopplungsmechanismen in der best-case Studie**

Laut der zuvor herausgearbeiteten Wirkung von Kopplungsmechanismen müsste in der Studie zu Kanada angenommen werden, dass aufgrund der vorbildlichen Ergebnisse in PISA eher ex ante und ex ongoing Kopplungsmechanismen eingesetzt werden als ex post Instrumente. Die Tatsache, dass in Alberta New Public Management eingeführt wurde, das heißt, de jure freie Schulwahl herrscht, die Schulen Haushaltskompetenzen inne haben und die Ergebnisse aus landesweiten Tests veröffentlicht werden, legt eher nahe, dass in Alberta ex post Kopplungsmechanismen vorzufinden sind. Allerdings hat Alberta zudem Professionelle Lerngemeinschaften<sup>15</sup> eingeführt. Dies wiederum spricht für die Annahme, dass ex ongoing Instrumente, eben Kooperation zwischen z.B. Lehrpersonen, eine große Rolle spielen. Aufgrund dieser sich teilweise widersprechenden Annahmen, wird in der Studie A explorativ untersucht, wie Governance in einem best-case funktioniert.

### *6.1 Fragestellungen der Studie A*

Die konkreten Forschungsfragen lauten:

- (1) Welche Elemente des NPM und der PLG werden in einem „best-case“ de facto umgesetzt?
- (2) Welche Probleme und Vorteile ergeben sich aus der Kombination der beiden Konzepte?

Hinsichtlich der ersten Frage wurde erwartet, dass Elemente des NPM und der PLG, die de jure implementiert sind, auch de facto zu beobachten sind. Allerdings wurde angenommen, dass Wettbewerbselemente des NPM, die in anderen Staaten negative Effekte

---

<sup>15</sup> Professionelle Lerngemeinschaften zeichnen sich durch Kollegialität (geteilte Normen, Zusammenarbeit, Mitbestimmung), Professionalisierung (Weiterbildung) und strukturelle wie soziale Unterstützung aus (vgl. Jäger, 2014, Tabelle 1).

erzeugen, eher in den Hintergrund treten, und dass stattdessen eher unterstützende Faktoren relevant sind, wie in der Literatur zu Steuerungsstrukturen vermutet wird (u.a. DIPF, 2007; Van Ackeren & Brauckmann, 2010).

Hinsichtlich der zweiten Frage zu den Problemen und Vorteilen der Kombination der Konzepte, wird explorativ analysiert, inwiefern die Spannungsverhältnisse zwischen NPM und PLG, die als Spannung zwischen Zentralisierung und Partizipation, Output- und Inputsteuerung sowie Wettbewerb und Kommunikation zusammengefasst werden können, aufgelöst werden bzw. wie mit ihnen umgegangen wird.

## 6.2 Methoden der Studie A

Der Staat Kanada sowie die Provinz Alberta wurden aufgrund ihrer Ergebnisse in PISA, die sie als positiven Sonderfall im OECD-Vergleich darstellen, ausgewählt (vgl. OECD, 2007; OECD, 2010). So erzielen die SchülerInnen in Alberta laut der als Grundlage dienenden Daten der OECD (2007) aus 2006 eine sehr gute Performanz und gleichzeitig zeigen sich im OECD-Vergleich geringe Disparitäten zwischen den SchülerInnenleistungen (OECD, 2007). Die Wahl der Schule innerhalb der Großstadt Edmonton erfolgte aufgrund von Aussagen von ExpertInnen aus dem Edmonton Public School Board. Sie empfahlen die untersuchte Schule als einen hinsichtlich der Sozialindikatoren in Edmonton und der Leistung der Provinz Alberta durchschnittlichen Fall.

Tabelle 2: Eingesetzte Methoden der Studie A pro Untersuchungsbereich

Methode	Merkmale der Schule	NPM	PLG
Dokumentenanalyse Standardisierter Fragebogen Quantitative Netzwerkanalyse	Durchgeführt Lehrpersonen SchülerInnen - Schulleitung Stellv. Schul-leitung	Durchgeführt Lehrpersonen - Schulleitung Stellv. Schulleitung LP_F Sozialarbeiterin Assistant Superintendent	Durchgeführt Lehrpersonen Lehrpersonen Schulleitung Stellv. Schulleitung LP_F SozialarbeiterIn Assistant Superintendent
Teilnehmende Beobachtung	-	Elterntreffen	-
ExpertInnen- diskussion	-	Interessierte des Research Support Services (Edmonton Public School Board)	-

Anmerkung: stellv. = stellvertretend; LP\_F = Lehrperson in spezieller Funktion

Um einen umfassenden Eindruck des untersuchten Falles zu erhalten, wurde eine Daten- und Methodentriangulation vorgenommen (vgl. u.a. Flick, 2008). Quantitative (Fragebögen, Netzwerkanalyse) und qualitative Erhebungsinstrumente (Interviews, Dokumentenanalyse) wurden im Jahr 2010 zur Analyse des Falles eingesetzt (vgl. Tabelle 2). Die quantitativen Daten wurden deskriptiv ausgewertet und die qualitativen Daten inhaltlich analysiert. Merkmale der Schule wurden mittels gängiger Indikatoren wie Bildungshintergrund der SchülerInnen etc. erhoben, Kriterien zu den beiden Steue-



rungskonzepten NPM und PLG wurden theoretisch erarbeitet und ihre Umsetzung wurde in Interviews bzw. in den Fragebögen erfragt.

Die Daten beruhen auf Aussagen der SchülerInnen aus dem Jahrgang der Stufe 9 (n = 194) sowie auf Aussagen des Lehrkollegiums (n = 30). Der Rücklauf bei den Lehrpersonen mit 73% ist als sehr zufriedenstellend, der für die SchülerInnen mit 40% als ausreichend einzuordnen. Insgesamt wurden an der Schule neun Interviews und eine teilnehmende Beobachtung bei einem Elterntreffen durchgeführt. Zudem wurden VertreterInnen des Bezirks interviewt. Dies sind der Assistant Superintendent, der/die SozialarbeiterIn und der/die SchulpsychologIn. Ein Gespräch mit dem „Trustee“ und dem Superintendenten konnte leider nicht realisiert werden. Im Rahmen eines Vortrags (12.02.2010) über das Projekt gewährten VertreterInnen des Research Support Services des Edmonton Public School Boards Einblick in ihre Arbeit.

### 6.3 Ergebnisse der Studie A

Allgemein lässt sich zum Kontext der Schule festhalten, dass die SchülerInnenleistung an der untersuchten Schule im Durchschnitt der provinzweiten Tests in Alberta liegt und somit, wie erwartet, wahrscheinlich oberhalb des OECD-Durchschnitts. Ob die Schule hinsichtlich der Sozialindikatoren Albertas im Durchschnitt liegt, kann nicht eingeschätzt werden. Bei der Schule handelt es sich um eine Schule mit einem Kollegium aus 30 Lehrpersonen und der familiäre Hintergrund der SchülerInnen ist eher bildungsnah (Aussage der Schulleitung und Befragung der SchülerInnen). Die SchülerInnen und Lehrpersonen nehmen als Probleme an ihrer Schule Konflikte des Miteinanders (Mobbing unter SchülerInnen) und den Schwierigkeitsgrad des Unterrichts wahr (standardisierte Befragung).

Die Ergebnisse zur ersten Frage der Studie A zeigt Tabelle 3. Hier sind Elemente des New Public Managements (NPM) wie freie Schulwahl und die Vorgabe und Überprüfung von Bildungsstandards sowie Elemente der Professionellen Lerngemeinschaften (PLG) wie Professionalisierung und Kollegialität aufgeführt. In der Tabelle sind alle Elemente des NPM und der PLG, die an der untersuchten Schule umgesetzt werden, fett gedruckt. Elemente, die nur teilweise praktiziert werden, sind kursiv gestellt, nicht im Fallbeispiel beobachtete Faktoren sind nicht hervorgehoben.

Hinsichtlich der Frage, welche Elemente des NPM und der PLG in einem „best-case“ de facto umgesetzt werden, bestätigt sich die Erwartung, dass in Edmonton Wettbewerbselemente weniger umgesetzt werden und stattdessen mehr auf Unterstützung und Kooperation gesetzt wird (vgl. DIPF, 2007; Van Ackeren & Brauckmann, 2010). So wird z.B. kein Wettbewerb zwischen den Lehrpersonen durch befristete Verträge forciert und auch die Schulen stehen nicht aufgrund der freien Schulwahl im Wettbewerb untereinander. Sie sind stattdessen zur Kooperation untereinander durch Netzwerke, die vom School Board initiiert wurden, aufgerufen. Auch erhalten die Schulen keine Boni oder Sanktionen, wenn ihre SchülerInnen gute bzw. schlechte Resultate z.B. in den provinzweiten Tests erhalten. Interessant ist zudem, dass zwar die Arbeit aller Beteiligten evaluiert und auch die Arbeit der Lehrpersonen im Unterricht beobachtet wird, es finden jedoch in der Regel keine negativen Sanktionen statt, sondern es wird bei Mängeln der Leistungserbringung Unterstützung angeboten.

Tabelle 3: Charakteristika von NPM und PLG

NPM	PLG
Voraussetzungen	
<ul style="list-style-type: none"> <li>• <i>Freie Schulwahl</i></li> <li>• <b>Rankings und Schulbeurteilungen</b></li> <li>• <i>Sanktionierbarkeit der Akteure seitens der Regierung</i></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Kollegialität (geteilte Normen, Zusammenarbeit, Mitbestimmung)</b></li> <li>• <b>Professionalisierung (Weiterbildung)</b></li> <li>• <b>Unterstützung (strukturell und sozial)</b></li> </ul>
Elemente der Zentralisierung	
<ul style="list-style-type: none"> <li>• <b>Definition von Bildungszielen seitens der Regierung</b></li> <li>• <b>Überprüfung von Bildungszielen seitens der Regierung</b></li> </ul>	
Elemente der Dezentralisierung	
<ul style="list-style-type: none"> <li>• Personalausonomie der Schule</li> <li>• <i>Sachmittelbewirtschaftung seitens der Schule</i></li> <li>• Lehrmittel sowie Unterrichtsgestaltung obliegen der Schule</li> <li>• <i>Beteiligung der Akteure der Schule an Entscheidungen</i></li> </ul>	

Anmerkung: Kursiv: wird in Ansätzen umgesetzt; fett gedruckt = wird praktiziert; keine Hervorhebung = wird nicht praktiziert; Quelle: Jäger, 2014.

Hinsichtlich der Umsetzung der PLG sind die Charakteristika Kollegialität, Professionalisierung und die Unterstützung der horizontalen Kommunikationsprozesse realisiert. Die vertikalen Einflussmöglichkeiten der Akteure sind indes durch die zentralen Elemente der Steuerung (Bildungszielvorgaben etc.) stark eingeschränkt. Die Qualität der Arbeit innerhalb der PLG kann darüber hinaus in dieser Studie nicht analysiert werden.

Die zweite Frage der Studie A möchte klären, welche Vorteile und Probleme sich aus der Kombination der beiden Konzepte ergeben. Hier wurden drei Spannungsverhältnisse zwischen den Konzepten NPM und PLG beleuchtet: Zentralisierung vs. Partizipation, Output- vs. Inputsteuerung, Wettbewerb vs. Kommunikation.

Der erst genannte Widerspruch zwischen Zentralisierung und Partizipation wird nicht vollkommen aufgelöst: So findet Partizipation zwar innerhalb von Teams und hierarchieübergreifenden Zirkeln statt, die Reichweite von inhaltlichen Entscheidungen scheint jedoch aufgrund der zentralen, hierarchischen Steuerungselemente sehr begrenzt zu sein. Das zweite Spannungsverhältnis ist hinsichtlich einer Input- und Prozesssteuerung entschieden. Hinzu kommt, dass das Output kontinuierlich gemessen wird, um das Input und die Prozesse zu optimieren. So werden die SchülerInnenleistungen durch Vergleichsarbeiten, kontinuierlich während der Schulzeit überprüft und es werden psychologische Tests durchgeführt; bei den Akteuren des Systems wird teilweise jährlich die Arbeit evaluiert und die Arbeit der Lehrpersonen wird durch Monitoring nahezu täglich begleitet. Zusätzlich ist hier interessant, dass z.B. das Monitoring von unterschiedlichen Akteuren übernommen wird, so wird Macht auf mehrere Zentren verteilt, die sich gegenseitig kontrollieren („checks and balances“, Strøm, 2000). Zudem steht die kontinuierliche Weiterbildung aller Akteure im Vordergrund, was einem Empowerment der Akteure entspricht.

Hinsichtlich des Spannungsverhältnisses zwischen Wettbewerb und vertrauensvoller Kommunikation zwischen den Akteuren, zeugen die Aussagen zu den Peer Netzwerken, Teams und z.B. der Rolle des Assistant Superintendent von einer vertrauensvollen Kommunikationskultur.

Als Hypothese für weitere Forschung listet *Tabelle 4* die Merkmalskombination des untersuchten „best-case“ der Bildungsperformanz, die als möglicherweise erfolgsversprechend abgeleitet wird. Da es sich um eine einzelne Fallstudie handelt, müssen die Ergebnisse quantitativ oder durch weitere Fallstudien validiert werden.

Tabelle 4: Erfolgversprechende Faktoren?

Ebene	Erfolgversprechende Faktoren der Governance im Bildungssystem?
Provinzebene	<ul style="list-style-type: none"> <li>• Vorgabe von Bildungszielen</li> </ul>
Lokale Verwaltungsebene	<ul style="list-style-type: none"> <li>• Zusätzliche demokratische Legitimation</li> <li>• Strukturelle und inhaltliche Unterstützung, besonders der Kommunikationsprozesse (PLG) in den Schulen und zwischen Verwaltung und Schule</li> </ul>
Schulebene	<ul style="list-style-type: none"> <li>• Teamarbeit (PLG)</li> </ul>
Ebenenübergreifende Merkmale	<ul style="list-style-type: none"> <li>• Professionalisierung aller Akteure</li> <li>• „checks and balances“-System</li> <li>• kontinuierliches Monitoring der Arbeit aller Akteure</li> <li>• Kopplungen der Ebenen durch PLG</li> <li>• Evaluationen ohne Sanktionen, feste Verträge zumindest für das Lehrpersonal</li> </ul>

Quelle: Jäger, 2014

## 7 Erwartete, nicht intendierte Effekte in den Studien zum Zentralabitur

In den Studien B, C und D wird angenommen, dass die Einführung des Zentralabiturs in bestimmten Bereichen der Schulqualität nicht intendierte Effekte, von denen einige zudem als nicht erwünscht eingeordnet werden, bewirkt. Weiter wird argumentiert, dass es zwar nicht intendierte Effekte aufgrund verschiedener Ursachen auftreten und ein Teil der nicht intendierten Effekte auch nicht eliminiert werden kann. Es wird aber angenommen, dass aufgrund der Einführung zentraler Prüfungen Kopplungsmechanismen verändert wurden, und diese Veränderung – neben intendierten erwünschten Effekten – auch systematisch nicht intendierte, unerwünschte Effekte hervorruft. So hat die Implementation des Zentralabiturs Kopplungsmechanismen in zwei Weisen verändert:

1. Das Verhältnis zwischen der Lehrperson im Unterricht und den Vorgaben des Ministeriums soll durch die ex post Kontrolle der zentralen Prüfung enger gekoppelt werden als zuvor als der Unterricht hauptsächlich über die Lehrperson mit der Prüfung verbunden war und die Lehrperson ex ante und ex ongoing, weltanschaulich mit den Vorgaben des Ministeriums verknüpft war. Somit soll die formal lose Kopplung im Unterricht zumindest zeitlich terminiert und the-

matisch begrenzt werden durch einen thematisch festgelegten und terminierten (Abiturthemen und Zeitraum von der zehnten bis zur zwölften Klasse), formalen ex post Kopplungsmechanismus. Ex post Kopplungsmechanismen können aber als Eingrenzung der Handlungsspielräume erlebt werden und einen Druck auf die Akteure erzeugen, der zu einer Übergeneralisierung und Demotivierung führt (Frey & Jegen, 2002).

2. Die Einführung der zentralen Prüfungen stellt zweitens eine Entkopplung zwischen Lehrperson und Prüfung dar. Dies bedeutet, dass hier nun eine lose Kopplung vorliegt, die von den Lehrpersonen verlangt, Interpretationsleistungen (hinsichtlich der Vorgaben zu den Abiturprüfungen) zu erbringen. Hier sich anschließende Fragen für die Lehrpersonen sind z.B.: Was wird im Abitur erwartet? Wie kann ich die SchülerInnen adäquat vorbereiten? Somit kann Unsicherheit entstehen, die zu nicht intendierten Effekten führen kann. Dieser Unsicherheit könnte beispielsweise durch zusätzliche Kopplungsmechanismen wie ex ante (Fortbildungen) oder aber auch ex ongoing Mechanismen (Handreichungen für den Unterricht, Kooperation zwischen Lehrpersonen) abgefedert werden.

In den Studien B, C und D wird damit angenommen, dass die Einführung eines zusätzlichen Kopplungsmechanismus in einem zuvor formal entkoppelten Bereich und die zeitgleiche Entkopplung eines zuvor eng gekoppelten Bereiches in einer SPDO von den Lehrenden zum Einen als Vertrauensbruch interpretiert wird und zum Zweiten zu Unsicherheit führt. Ersteres kann die für die SPDP notwendige intrinsische Motivation verdrängen (Frey & Jegen, 2000). Beides kann nicht intendierte Effekte, und zwar auch unerwünschte, zur Folge haben.

### *7.1 Fragestellungen der Studien B, C und D*

In den drei Studien wurde die Frage untersucht, inwiefern die Implementation des Zentralabiturs in Hessen und Bremen nicht intendierte Effekte zur Folge hat und zwar in den folgenden Bereichen, die für die Schulqualität relevant sind:

- Kongruenz zwischen Unterricht und Abiturprüfung
- Eingrenzung der Themenvarianz<sup>16</sup> (Teaching to the test)
- Arbeitsunzufriedenheit
- individuelle und kollektive Selbstwirksamkeit
- Schulklima

Im Rahmen der Studien wurde nicht untersucht, inwiefern bestimmte Effekte intendiert, vorhersehbar oder beherrschbar sind (Braun, 2008). Es wurde auch nicht untersucht, welche Effekte (un-)erwünscht sind. Stattdessen wurden anhand der Literatur folgende Effekte als unerwünscht definiert:

- eine variierende Inkongruenz zwischen Unterricht und Prüfung, da dies in der Vorbereitung auf das Abitur ein Wettbewerbsnachteil für die betroffenen Schü-

---

<sup>16</sup> Themenvarianz wird hier u.a. verstanden als das Ausmaß, in dem die Interessen der Schüler/-innen und aktuelle Themen im Sachfach bei der inhaltlichen Auswahl der Unterrichtsthemen berücksichtigt werden.

lerInnen bedeutet. Hier wird angenommen, dass die Lehrpersonen Kongruenz herstellen werden können, dass dies aber nicht allen Lehrpersonen zu gleichem Maße gelingt.

- ein Teaching to the test, da dies einem anregenden Unterricht zuwider läuft (vgl. zu Formen von Teaching to the test: Au, 2007). Hier wird die Hypothese aufgestellt, dass trotz low-stakes System ein Teaching to the test auftritt, da das ex post Instrument zu einem Verdrängungseffekt führt (Frey & Jegen, 2002).
- eine Steigerung der Arbeitsunzufriedenheit, da Zufriedenheit mit der Arbeit wichtig für das soziale Leben, aber auch für die Motivation der Lehrenden ist (Fend, 1998, S. 104). Hier wird angenommen, dass die Arbeitsunzufriedenheit steigt u.a. aufgrund der Eingrenzung des Handlungsspielraums durch das Zentralabitur (Berryhill, Linney & Fromewick, 2009).
- eine sinkende individuelle wie kollektive Selbstwirksamkeit im Zusammenhang mit der Implementierung zentraler Prüfungen, da dies dazu führen kann, dass schwierige Aufgaben gemieden werden (Schmitz & Schwarzer, 2000), was sich negativ auf den Lehrberuf mit seinen ständig neuen Situationen und Aufgaben auswirkt. Auch hier wird angenommen, dass die Einführung des Zentralabiturs als Eingrenzung von Handlungsspielräumen verstanden wird und es zu einer Verringerung der Wahrnehmung von individueller und kollektiver Selbstwirksamkeit kommt (Brockmeyer & Edelstein, 1997, S.13).

Ein Effekt auf das Schulklima wurde dagegen nicht erwartet, da das Schulklima in dieser Untersuchung vornehmlich über soziale Faktoren gemessen wurde und hier kein Zusammenhang mit den zentralen Prüfungen erwartet wird.

Zusätzlich werden in den Studien neben der Wirkung des Zentralabiturs weitere Einflussfaktoren untersucht. An dieser Stelle werden nur die Fragen und Ergebnisse, die in Bezug auf die Kopplungsmechanismen interessant sind, berichtet. So wird in Studie B unter anderem gefragt, ob die Kooperation zwischen den Lehrpersonen zum Curriculum den Teaching to the test Effekt abschwächt. Es wird sich hier ein positiver Einfluss auf die Themenvarianz versprochen, also eine Abschwächung des Teaching to the test, da die Kooperation zum Curriculum eine weitere Kopplung, eine relationale, darstellt. Die Tabelle 5 zeigt die untersuchten Fragestellungen pro Studie im Überblick.

## *7.2 Methoden der Studien B, C und D*

Die Daten der befragten Lehrpersonen und SchülerInnen in Bremen und Hessen lagen zum Zeitpunkt der Berechnungen für die Artikel für unterschiedliche Jahre vor: So basiert Studie B auf den Daten zu 2007 und 2008 und die Studien C und D konnten zusätzlich die Daten aus 2009 berücksichtigen. Studie C konzentriert sich auf die Analyse der Aussagen der Lehrpersonen, wohingegen Studie B und D auch Aussagen der SchülerInnen miteinbeziehen.

Für detaillierte Angaben zum Rücklauf wird an dieser Stelle auf die Artikel verwiesen. Hier wird als Anhaltspunkt die Datengrundlage der gesamten Studie für Lehrpersonen und SchülerInnen in Tabelle 6 und Tabelle 7 dargelegt (vgl. Maag Merki & Oerke, 2012).

Tabelle 5: Fragestellungen der Studien B, C und D im Überblick

Studie B*		Studie C**		Studie D***	
1	Führen zentrale Abiturprüfungen unter low-stakes <sup>17</sup> Bedingungen zu einem Teaching to the test?	1	Zeichnet sich ein Teaching-to-the-test-Effekt ab und verändert sich dieser in Abhängigkeit der Erfahrung mit dem Zentralabitur?	1	Wie werden die Aspekte Schulklima, Arbeitszufriedenheit und individuelle sowie kollektive Selbstwirksamkeit von den Lehrpersonen über die Jahre wahrgenommen? Wie nehmen Schüler/-innen das Schulklima wahr?
2	Beeinflussen die Faktoren Unsicherheit gegenüber den zentralen Prüfungen, die Erfahrung mit den zentralen Prüfungen, Kooperation unter den Lehrpersonen hinsichtlich des Curriculums und die individuelle Selbstwirksamkeit die Eingrenzung der Themenvarianz?	2	Welche Faktoren beeinflussen die Verwirklichung von Themenvarianz im Unterricht der zentral geprüften Fächer?	2	Zeigt sich ein Zusammenhang zwischen der Erfahrung der Lehrpersonen mit dem zentralen Prüfungssystem und den genannten Konstrukten?
		3	Wie beurteilen die Lehrpersonen nach dem Abitur die Übereinstimmung zwischen der Vorbereitung im Unterricht und den zentralen Prüfungen?	3	Stehen persönliche und biografische Faktoren im Zusammenhang mit den untersuchten Konstrukten?
		4	Verbessert sich durch die Erfahrung mit den zentralen Prüfungen die Kongruenz zwischen Unterricht und Prüfung aus der Perspektive der Lehrpersonen?		
		5	Gibt es Faktoren, die die Wahrnehmung der Kongruenz zwischen Unterricht und zentralen Prüfungen aus Sicht der Lehrpersonen erhöhen bzw. konterkarieren?		

Anmerkung: \*, „State-wide, low-stakes tests and a teaching-to-the-test effect? An analysis of data from two German Bundesländern“ (Jäger, Maag Merki, Oerke & Holmeier, 2012); \*\*, „Herausforderung Zentralabitur: Unterrichtsinhalte an Prüfungsthemen anpassen, ohne dabei die Interessen der Schüler/-innen zu vernachlässigen“ (Jäger 2012a); \*\*\*, „Schulklima, Selbstwirksamkeit und Arbeitszufriedenheit aus Sicht der Lehrpersonen und Schüler/-innen in Hessen und Bremen“ (Jäger 2012b)

Tabelle 6: Stichprobe und Rücklauf der Lehrpersonen

Jahr	Stichprobe vor dem Abitur (N, Rücklauf) Population 1*			Stichprobe nach dem Abitur (N, Rücklauf) Population 2**		
	2007	2008	2009	2007	2008	2009
Hessen	N = 1093 72%	N = 949 59%	N = 1005 59%	N = 529 67%	N = 430 59%	N = 479 48%
Bremen	N = 641 65%	N = 577 60%	N = 424 45%	N = 213 56%	N = 298 55%	N = 230 37%
Gesamt	N = 1734 69%	N = 1526 59%	N = 1429 54%	N = 742 63%	N = 728 57%	N = 709 44%

Anmerkung: \*Alle Lehrpersonen in den Schulen; \*\*Lehrpersonen der Jahrgangsstufe 13 mit Beteiligung am Zentralabitur: Unterricht in einem zentral geprüften Kurs, in dem Schüler/-innen das Abitur schrieben, und/oder Zweitkorrektur von Abituraufgaben; Quelle: Maag Merki & Oerke, 2012, S. 56

<sup>17</sup> Low-stakes, vgl. Fußnote 8

Die abhängige Variable in der Studie B ist die Eingrenzung der Themenvarianz, in Studie C ist die abhängige Variable einmal die Eingrenzung der Themenvarianz und einmal die Kongruenz zwischen der Vorbereitung im Unterricht und den zentralen Prüfungen. In der Studie D wurden als abhängige Variablen das Schulklima, die Arbeitsunzufriedenheit sowie die individuelle und kollektive Selbstwirksamkeit analysiert. Je nach Themenbereich wurden unterschiedliche Einflussfaktoren untersucht (unabhängige Variablen) (siehe unten). Hinsichtlich der unabhängigen Variablen werden aus Platzgründen nur die Ergebnisse berichtet, die für die Kopplungsmechanismen relevant sind. Dies sind Variablen zur Unsicherheit und zur Kooperation, denn erstere erfordern als Ausgleich Kopplungsmechanismen und letztere koppeln relational.

Die Auswertungen der Daten erfolgten zuerst deskriptiv, wobei Differenzen bei den Lehrpersonen mittels eines t-Tests untersucht werden. Bei den SchülerInnen werden Unterschiede zwischen den Jahren innerhalb eines Bundeslandes durch eine Regressionsanalyse berechnet. Zudem werden die Daten der Lehrpersonen (und teilweise der SchülerInnen) mittels des Softwareprogramms „Hierarchical Linear and Nonlinear Modeling“ (HLM 6.06) von Raudenbush, Bryk und Congdon (2004) unter Berücksichtigung der Mehrebenenstruktur analysiert. Hier werden Zwei- und Dreiebenenmodelle analysiert. Dreiebenenmodelle werden hier durchgeführt, sobald mehrere Jahre untersucht werden. Im Zweiebenenmodell werden auf der ersten Ebene variierende Faktoren innerhalb der Individuen erfasst und auf der zweiten Ebene Unterschiede zwischen Individuen. Im Dreiebenenmodell beinhaltet die erste Ebene die intraindividuellen Veränderungen über die Jahre, die zweite Ebene die Unterschiede zwischen den Individuen (z.B. Geschlecht), die dritte Ebene ergänzt sodann die Differenzen zwischen den Schulen (z.B. das Bundesland). Fehlende Werte stellten kein Problem dar, da für die erste Ebene in Mehrebenenanalysen keine gleiche Anzahl von Beobachtungen vorausgesetzt wird (Hox, 2002, S. 95f).

Tabelle 7: Stichprobe und Rücklauf der SchülerInnen

Bundesland	Jahr	Stichprobe Fragebogen vor dem Abitur (Rücklauf)	Stichprobe Fragebogen nach dem Abitur (Rücklauf)
Hessen	2007	N = 973 (68%)	N = 826 (57%)
	2008	N = 975 (71%)	N = 793 (58%)
	2009	N = 1017 (74%)	N = 876 (64%)
Bremen	2007	N = 751 (50%)	N = 570 (38%)
	2008	N = 977 (65%)	N = 479 (32%)
	2009	N = 1012 (71%)	N = 481 (34%)
Gesamt	2007	N = 1724 (58%)	N = 1396 (47%)
	2008	N = 1952 (68%)	N = 1272 (44%)
	2009	N = 2029 (72%)	N = 1357 (48%)

Quelle: Maag Merki & Oerke, 2012, S. 58

So wurden die folgenden Regressionsgleichungen in den Studien eingesetzt. „ZA“ steht für das Zentralabitur, „KFT“ für den Kognitiven Fähigkeitstest, die Variablen „ErfahrungZA“ und „BeteiligungZA“ unterscheiden sich insofern, dass erstere eine Aussage dazu macht, wie oft eine Lehrperson an zentralen Prüfungen bis zum Erhe-

bungszeitpunkt beteiligt war (Erfahrung mit den zentralen Prüfungen), die Variable „BeteiligungZA“ dagegen informiert, ob eine Lehrperson im laufenden Jahr von zentralen Prüfungen betroffen ist. In den Regressionsgleichungen sind die Variablen kursiv gedruckt, die zentriert (grand mean) in das Modell aufgenommen werden:

Studie B:

- Themenvarianz<sub>2008</sub>, zentral und dezentral geprüfte Kurse =  $\gamma_{00} + \gamma_{01} * \text{Bundesland} + \gamma_{10} * \text{ErfahrungZA} + \gamma_{20} * \text{BeteiligungZA} + u_0 + r$
- Themenvarianz<sub>2008</sub>, zentral geprüfte Kurse =  $\gamma_{00} + \gamma_{01} * \text{KFT Median} + \gamma_{02} * \text{KFT Standardabweichung} + \gamma_{03} * \text{Bundesland} + \gamma_{10} * \text{Kooperation Curriculum} + \gamma_{20} * \text{Individuelle Selbstwirksamkeit} + \gamma_{30} * \text{UnsicherheitZA} + \gamma_{40} * \text{ErfahrungZA} + u_0 + r$

Studie C:

- Themenvarianz<sub>2007-2009</sub>, zentral und dezentral geprüfte Kurse =  $\gamma_{000} + \gamma_{001} * \text{Bundesland} + \gamma_{010} * \text{Geschlecht} + \gamma_{100} * \text{BeteiligungZA} + \gamma_{200} * \text{ErfahrungZA} + \gamma_{300} * \text{Jahr} + r_0 + u_{00} + e$
- Themenvarianz<sub>2007-2009</sub>, zentral geprüfte Kurse =  $\gamma_{000} + \gamma_{001} * \text{Bundesland} + \gamma_{010} * \text{Geschlecht} + \gamma_{100} * \text{BeteiligungZA} + \gamma_{200} * \text{ErfahrungZA} + \gamma_{300} * \text{Jahr} + \gamma_{400} * \text{Kollektive Selbstwirksamkeit} + \gamma_{500} * \text{UnsicherheitZA} + \gamma_{600} * \text{Kooperation Curriculum} + r_0 + u_{00} + e$
- Kongruenz<sub>2007-2009</sub> =  $\gamma_{000} + \gamma_{001} * \text{Bundesland} + \gamma_{010} * \text{Geschlecht} + \gamma_{100} * \text{ErfahrungZA} + \gamma_{200} * \text{Jahr} + \gamma_{300} * \text{UnsicherheitZA} + \gamma_{400} * \text{Kooperation Curriculum} + \gamma_{500} * \text{Kollektive Selbstwirksamkeit} + \gamma_{600} * \text{Themen}^{18} + \gamma_{700} * \text{Fortbildung}^{19} + \gamma_{800} * \text{Operatoren}^{20} + r_0 + u_{00} + e$

Studie D:

- Schulklima/ Arbeitsunzufriedenheit/ Individuelle Selbstwirksamkeit /Kollektive Selbstwirksamkeit<sub>2007-2009</sub> =  $\gamma_{000} + \gamma_{001} * \text{Bremen} + \gamma_{010} * \text{Geschlecht} + \gamma_{100} * \text{ErfZentral} + \gamma_{200} * \text{Jahr} + \gamma_{300} * \text{Schuldienst_k} + \gamma_{400} * \text{Schuldienst_l} + r_0 + u_{00} + e$
- Schulklima/ Arbeitsunzufriedenheit/Kollektive Selbstwirksamkeit<sub>2007-2008, 2008-2009</sub> =  $\gamma_{000} + \gamma_{001} * \text{Bremen} + \gamma_{010} * \text{Geschlecht} + \gamma_{100} * \text{ErfZentral} + \gamma_{200} * \text{Jahr07} + \gamma_{300} * \text{Jahr09} + \gamma_{400} * \text{Schuldienst_k} + \gamma_{500} * \text{Schuldienst_l} + r_0 + u_{00} + e$

Die Tabelle 8 zeigt die methodischen Differenzen im Überblick.

<sup>18</sup> Einzelitem: „Hilfreiche Grundlagen für die angemessene Vorbereitung auf die zentralen Abiturprüfungen waren die Angaben zu den verbindlichen Schwerpunktthemen für das Abitur.“

<sup>19</sup> Einzelitem: „Hilfreiche Grundlagen für die angemessene Vorbereitung auf die zentralen Abiturprüfungen waren die Fortbildungsveranstaltungen zu den Schwerpunktthemen.“

<sup>20</sup> Einzelitem: „Hilfreiche Grundlagen für die angemessene Vorbereitung auf die zentralen Abiturprüfungen war die Liste der Operatoren.“



Tabelle 8: Methodische Differenzen zwischen den Studien B, C und D

	Studie B*	Studie C**	Studie D***
Jahr	2007, 2008	2007, 2008, 2009	2007, 2008, 2009
Anzahl der Ebenen in der Mehrebenenanalyse	2008: Zwei Zweiebenen-Modelle a) zentral und dezentral geprüfte Kurse gemeinsam b) nur zentral geprüfte Kurse	Zwei Dreiebenenmodelle zu Themenvarianz a) zentral und dezentral geprüfte Kurse gemeinsam b) nur zentral geprüfte Kurse Ein Dreiebenenmodell zur Kongruenz	Zwei Dreiebenenmodelle (siehe Effekte über die Zeit) zu Schulklima, Arbeitsunzufriedenheit, und kollektiver Selbstwirksamkeit Ein Dreiebenenmodell zu individueller Selbstwirksamkeit <sup>†</sup>
Effekte über die Zeit	-	lineare Effekte	a) lineare Effekte von 2007 zu 2009 b) quadratische bzw. U-förmige Effekte von 2007 zu 2008 und 2008 zu 2009
Datengrundlage	Lehrpersonen SchülerInnen	Lehrpersonen	Lehrpersonen SchülerInnen

Anmerkung: \* Jäger, Maag Merki, Oerke & Holmeier, 2012; \*\* Jäger, 2012a; \*\*\* Jäger, 2012b; <sup>†</sup> Zu der Variable individuelle Selbstwirksamkeit liegen nur Daten für 2008 und 2009 vor, daher wurden nur diese analysiert.

### 7.3 Ergebnisse der Studien B, C und D

Die Ergebnisse der Studien B, C und D bestätigen weitgehend die dort formulierten Erwartungen. So erzielen die Lehrenden – aufgrund der Unterstützungsmaßnahmen und trotz Unsicherheit gegenüber den zentralen Prüfungen – eine Anpassung ihres Unterrichts an die abgefragten Inhalte im Abitur (Kongruenz, Studie C). Allerdings zeigt sich durch die Erfahrung mit den zentralen Prüfungen keine Steigerung der Kongruenz zwischen Unterricht und Abiturprüfung. Dies wurde auch nicht erwartet, da die Prüfungsthemen jedes Jahr neu sind und die Erfahrungen des Vorjahres seitens der Lehrpersonen nicht genutzt werden können. Die Ergebnisse zeigen sogar einen leicht negativen Effekt zwischen der Erfahrung mit den zentralen Prüfungen und der Kongruenz (Studie C). Dieses Ergebnis müsste durch z.B. fachspezifische Analysen weitergehend analysiert werden, um eventuell einer sich entwickelnden Inkongruenz entgegenwirken zu können.

Studie C zeigt zudem, dass die Kongruenz zwischen Prüfung und Vorbereitung im Unterricht, nicht in einem positiven Zusammenhang mit der Kooperation zwischen den Lehrpersonen zum Curriculum steht. Hilfestellungen seitens des Ministeriums dagegen stehen in einem positiven Zusammenhang mit der Kongruenz zwischen Unterricht und Prüfung.

Die Studien B und C belegen, dass trotz eines low-stake System ein Teaching to the test stattfindet. Studie B identifiziert zudem Faktoren, die mit einem Teaching to the test in den zentral geprüften Kursen einhergehen: So steht hohe Unsicherheit gegenüber den zentralen Prüfungen mit einer stärkeren Eingrenzung der Themenvarianz in einem

Zusammenhang, weitere Studien müssten zeigen, ob und wie sich die Unsicherheit über die Jahre verändert und wie sich dies dann auf die Eingrenzung der Themenvarianz auswirkt. Zudem zeigt sich, dass die Erfahrung mit den zentralen Prüfungen nicht dazu führt, dass Themen weniger eingegrenzt werden. Im Gegenteil: Lehrpersonen mit mehr Erfahrung mit dem zentralen Prüfungssystem grenzen die Themenvarianz mehr ein als Lehrpersonen mit weniger Erfahrung.

Die Wahrnehmung der eigenen, individuellen Selbstwirksamkeit steht dagegen mit der Themenvarianz in einem positiven Zusammenhang: Lehrpersonen, die sich als selbstwirksam erleben, grenzen die Themen weniger ein als Lehrpersonen, die ihre Selbstwirksamkeit geringer einschätzen.

Hinsichtlich der Variable Kooperation zum Curriculum zeigt die Studie B – entgegen der Erwartung –, dass Kooperation zum Curriculum in keinem Zusammenhang mit der Themenvarianz steht. Die Studie C dagegen, die nicht nur das Jahr 2008, sondern auch die Jahre 2007 und 2009 berücksichtigt, zeigt – erwartungskonform – einen signifikanten positiven Zusammenhang. Hier müssten weitere Analysen folgen, die z.B. die Entwicklung des Kooperationsverhaltens hinsichtlich dezentral und zentral geprüfter Fächer im Detail untersuchen. Auch wären detaillierte Studien zu dem Zusammenspiel von Kooperation, Unsicherheit und Kongruenz, z.B. anhand von Pfadmodellen aussichtsreich. Auf der Schulebene kann nur ein geringer Anteil an Varianz erklärt werden. Ein Effekt zeigt sich bei der Zugehörigkeit zu einem Bundesland, wobei die Ursache an dieser Stelle nicht geklärt werden kann.

Die Annahmen aus Studie D bestätigen sich: So ist hinsichtlich der Arbeitsunzufriedenheit ein schwacher Unzufriedenheitsanstieg im Zusammenhang mit der Erfahrung mit den zentralen Prüfungen zu beobachten. Dies mag darauf zurückzuführen sein, dass das Zentralabitur Umstellungen erfordert, die als Belastung erlebt werden können. Außerdem engt das Zentralabitur Handlungsspielräume der Lehrpersonen ein, was ebenfalls negativ empfunden werden kann. Schließlich könnte dies auch damit zusammenhängen, dass die Lehrpersonen ihren Status in der Gesellschaft als niedriger einschätzen als zu Zeiten des dezentralen Prüfungssystems und daher auch unzufriedener mit ihrer Arbeit sind (vgl. Bishop, 1999).

Die Erfahrung mit den zentralen Prüfungen mindert zudem die Wahrnehmung individueller und kollektiver Selbstwirksamkeit. Es zeigt sich ein leichter, signifikanter, negativer Zusammenhang. Dieser ist zu erwarten, da u.a. durch die eingeführte Unsicherheit gegenüber den Prüfungsfragen im Abitur, den Lehrpersonen Handlungs- und Planungssicherheit genommen wird.

Ein Effekt auf das Schulklima trat auch nicht auf. Dies entspricht ebenfalls der Erwartung, da das Schulklima hauptsächlich über soziale Aspekte gemessen wurde, welche von Reformen des Prüfungssystems im Abitur unberührt bleiben.

Hinsichtlich der Studie D muss kritisch angemerkt werden, dass alle untersuchten Einflussfaktoren, also auch die Einführung der zentralen Prüfungen, nur geringe Schwankungen erklären. In Bezug auf das Schulklima und die Arbeitsunzufriedenheit kann dies in der insgesamt nur geringen aufzuklärenden Varianz begründet sein. Hinsichtlich der Selbstwirksamkeitserwartungen liegt die Ursache aber auch darin, dass relevante Einflussfaktoren im Rahmen dieser Studie nicht analysiert werden (z.B. Er-

fahrung von Selbstwirksamkeit im täglichen Arbeitskontext, Verhalten der Schulleitung, etc.).

## 8 Diskussion

Governance-Strukturen ermöglichen und begrenzen Handlungen und sind daher für die Erzeugung von Leistungen relevant. Inwiefern bestimmte Koordinationsformen zwischen Akteuren im Mehrebenensystem die gewünschte Leistung verbessern können, ist in Bezug auf das Bildungssystem die übergeordnete Frage dieser Arbeit. Steuerung und Governance bezeichnen hier unterschiedliche Perspektiven auf „absichtsvolle Regelungen“ von (kollektiven) Akteuren. Knapp zusammengefasst, betrachtet die Steuerungstheorie das Handeln der Akteure und die Governance-Forschung die gesamte Struktur (vgl. Mayntz, 2008; Schuppert, 2008). Beiden Perspektiven bzw. Ansätzen ist gemeinsam, dass sie die Bereitstellung öffentlicher Güter analysieren.

Im Fokus dieser Arbeit stehen die Kopplungsmechanismen. Sie werden als Teil der Regelungsstrukturen der Governance verstanden und umfassen alle Arten der Verknüpfung zwischen (kollektiven) Akteuren. In der vorliegenden Arbeit wird für eine differenzierte Betrachtung von Kopplungsmechanismen plädiert, da so Wirkungszusammenhänge herausgearbeitet werden können.

So werden in der Arbeit Typologien von Manow (2005) und Lasky, Datnow und Stringfield (2005) sowie eine Differenzierung von Kopplungen von Richter (2012) vorgestellt. Kopplungsmechanismen können demnach darin unterschieden werden, ob sie vor (ex ante), während (ex ongoing) oder nach der Leistungserbringung (ex post) erfolgen, ob sie thematisch oder temporär begrenzt sind und ob sie formal, informell, strukturell, relational und/oder weltanschaulich (kollektive) Akteure koppeln. Diese Differenzierung erlaubt es, Alternativen zu erarbeiten: Denkt man nur an das Beispiel zentrale Prüfungen und die Möglichkeiten sie entweder ex ante, ex ongoing oder ex post einzusetzen, also als Eingangsprüfung, z.B. zum Gymnasium, als Prüfung während der Gymnasialzeit oder aber als zentrale Abschlussprüfung, wie das Zentralabitur, so zeigt sich, dass kleine Varianzen der Kopplungsmechanik unterschiedliche Wirkungen erzielen können. Daher ist die Annahme hier, dass durch Analysen von Kopplungsmechanismen Anregungen für die Gestaltung von Steuerung gewonnen werden können; zum Einen in Hinsicht auf die Erreichung intendierter Ziele, zum Anderen hinsichtlich der Verminderung nicht intendierter, unerwünschter Effekte.

Die Literatur zu Kopplungsmechanismen legt darüber hinaus nahe, Kopplungen kontext- bzw. situationsspezifisch zu analysieren (Lasky, Datnow & Stringfield, 2005), daher wurde in der vorliegenden Arbeit in einem nächsten Schritt das Bildungssystem als „soziale personenbezogene Dienstleistungsorganisation“ (SPDO) reflektiert (Drepper, 2010, S. 143). So liegen die Besonderheiten dieser Organisationsform darin, dass Personen das zu beeinflussende „Rohmaterial“ sind, dass die Ziele des Systems oftmals mehrdeutig und interpretativ umkämpft sind, die Organisation in einer vielschichtigen Umwelt agiert, Interaktion das Kerngeschäft darstellt und damit eine Technologievagheit wie auch ein Effizienzverlust einhergeht (vgl. Drepper, 2010, S.

143). Die Schule wurde dementsprechend als eine Handlungseinheit der SPDO begriffen, die ihre „Identität im Spannungsfeld und in der Auseinandersetzung mit ihrer staatlichen und professionellen Rahmenordnung“ definieren muss (Drepper & Tacke, 2012, S. 231).

Theoretisch, so wurde herausgearbeitet, sollte im Bildungssystem eine Kombination der Kopplungsmechanismen stattfinden: Einerseits sind Bildungssysteme aufgrund ihrer Eigenschaft als SPDO auf die intrinsische Motivation ihrer Mitarbeitenden angewiesen und sollten die Auswahl der Mitarbeitenden im Sinne der eigenen Ziele sicherstellen (ex ante Kopplungsmechanismus) und auf ex post Instrumente, die diese Motivation verdrängen könnten, weitgehend verzichten (vgl. Frey & Jegen, 2002). Zudem ist – ebenfalls aufgrund der Struktur der SPDOs – zu erwarten, dass nicht intendierte Effekte im Bildungssystem auftreten. Daher sollten ex ongoing Instrumente zur Fehlerquellenanalyse genutzt werden. Schließlich wurde in dieser Arbeit mit Döhler (2008) und Manow (2008) angenommen, dass auch abweichendes Verhalten bei Mitarbeitenden – als Ausnahme von der Regel – auftreten kann. Dieses Verhalten kann durch ex post Instrumente sanktioniert werden. Theoretisch wurde daher für eine Kombination der Instrumente argumentiert.

Die Forschung hat im Bildungssystem zwei empirisch zu beobachtende Kopplungsmechanismen herausgearbeitet: Zum einen, dass Kopplungsmechanismen im Bildungssystem legitimiert sein, das heißt, dass ihre Zielsetzungen einen Bezug zur Gesellschaft aufweisen müssen. Zum Zweiten, dass nur gering formal gekoppelt wird, da durch Entkopplung Spielräume bereit gestellt werden, um so u.a. die komplexen Anforderungen und die Situativität des Unterrichts adäquat bewältigen zu können. In den letzten Jahrzehnten wurden diese Kopplungsmechanismen jedoch reformiert: So wurden zusätzlich ex post Instrumente, wie das Zentralabitur eingeführt und die ex ongoing Instrumente verstärkt (Hospitationen und Inspektionen). Folgende Forschungsfragen warf die Synopse daher auf, für die die Studien Teilergebnisse liefern:

- I. Welche Kopplungsmechanismen können dazu beitragen, dass in Bildungssystemen Ziele umgesetzt werden, das heißt, die erwünschte Leistung erbracht wird?
- II. Wie wirken sich Reformen der Kopplungsmechanismen aus?
  - a. Welche positiven Effekte können erzielt werden?
  - b. Welche negativen Effekte ergeben sich?

Die beiden Fragen wurden in der vorliegenden Arbeit folgendermaßen spezifiziert:

1. Mittels welcher Kopplungsmechanismen wird in einem best-case gesteuert? (Studie A)
2. Erzeugen ex post Kopplungsmechanismen nicht intendierte Effekte? (Studie B, C und D)

### *8.1 Kopplungsmechanismen in einem best-case*

Die übergeordnete Frage der Fallstudie aus Alberta lautet: „Mittels welcher Kopplungsmechanismen wird in einem best-case gesteuert?“ Die Ergebnisse zeigen, dass in diesem best-case alle Möglichkeiten der Kopplung (formal, relational, strukturell, in-

formell und weltanschaulich) ausgeschöpft werden. Auch ist ein Schwerpunkt auf ex ante und ex ongoing Kopplungsmechanismen zu beobachten. Einige Beispiele:

1. Ex ante Kopplungsmechanismen:

- Das Personal der Schule sowie der Schulverwaltung wird ex ante durch Karriereverläufe ausgewählt, z.B. Schulleitung wird nur jemand, der/die zuvor Lehrperson war.
- Doppelte demokratische Absicherung: Die Werthhaftigkeit der Aufgabe wird durch eine lange Delegationskette von der Provinz zur Schule erzeugt, die auf der Ebene des Bezirks zusätzlich demokratisch abgesichert wird. Dies stellt somit die Integration der bezirksweiten Interessen in die Interessen der Provinz sicher. Dies Delegationskette ist hierarchisch (formaler, ex ante Kopplungsmechanismus), die Provinz wie das School Board formulieren Ziele (formal, ex ante) und stellen Finanzen bereit (struktureller, ex ante Kopplungsmechanismus) und schließlich können die WählerInnen ex post Kontrolle ausüben (vgl. Strøm, 2000).

2. Ex ongoing Kopplungsmechanismen:

- Der Mechanismus „compliance“ wird systemweit und auf der Ebene der Schule angewendet: Die Schule als Akteur formuliert Schulziele, die Akteure im System formulieren (zumeist) jährlich Zielvereinbarungen für sich selbst.
- Gremienverflechtungen existieren (z.B. ist jemand Teil des administrativen Teams und JahrgangsstufenkoordinatorIn).
- Es existiert ein System aus „checks and balances“ (Strøm, 2000).
- Monitoring auf allen Ebenen: SchülerInnenleistung, Arbeit der Lehrpersonen und der Schulleitung und weiterer Akteure.
- Auf der Ebene des School Boards finden Hearings für die Öffentlichkeit statt.
- Peer Netzwerke koppeln relational (unterstützen Kollegialität)
- Professionalisierung koppelt relational.

In Alberta kommen neben diesen Kopplungsmechanismen aber auch ex post Mechanismen zum Einsatz, die in der Theorie hinsichtlich der komplexen Aufgaben der Akteure als eher ungeeignet eingeordnet wurden. Die Rede ist hier von den befristeten Verträgen und den Evaluationen der Arbeit fast aller Akteure (abgesehen der Lehrpersonen). Interessanterweise zeigen die Interviews mit den befragten Akteuren jedoch, dass keine Sanktionen im Sinne der ex post Mechanismen erfolgen. Das bedeutet, dass das ex post Kontrollinstrument eher als eine „glaubwürdige Drohung“ (Osborne, 2004) eingeordnet werden muss. Wobei letztlich nicht klar ist, ob z.B. die Beendigung des Arbeitsverhältnisses als Sanktionsform besteht, also ob sie im Ernstfall angewendet wird oder nicht. Eine theoriekonforme Interpretation wäre, dass in Alberta eine konstruktive Fehlerkultur hinsichtlich nicht intendierter Effekte geübt wird. Die „glaubwürdige Drohung“ wäre in diesem Rahmen die letzte zu nutzende Möglichkeit

bei Fehlverhalten, um das System vor „shirking“ zu schützen. Im Rahmen dieser Studie konnte dies nicht näher untersucht werden.

Weitere Varianten den Einsatz der ex post Kontrolle zu erklären, sind: a) Die Evaluationen könnten darauf abzielen, den Wettbewerb zwischen den Akteuren zu stärken, um so extrinsische Motivation zu erhöhen. b) Die Evaluationen wurden nur auf politischen Druck hin eingeführt, sie sind somit nur eine strategische Antwort auf gesellschaftliche Erwartungen, de facto werden sie ignoriert. c) Sie zeugen von einem Veränderungsprozess des Systems, sie sind damit entweder Überbleibsel einer alten Struktur oder Beginn einer neuen. Alle diese Interpretationen sind indes nur Vermutungen, denen nachzugehen interessant ist.

Insgesamt kann damit festgehalten werden, dass in dem best-case hauptsächlich ex ante und ex ongoing Instrumente als Kopplungsmechanismen zur Steuerung eingesetzt werden und darüber hinaus versucht wird, durch verschiedene Arten zu koppeln, also multiplex zu koppeln. Es kommen zudem ex post Instrumente zum Einsatz, deren Bedeutung im Gesamtarrangement abschließend nicht geklärt werden kann („glaubwürdige Drohung“, Osborne, 2004, oder ist die Drohung nur ein Papiertiger?). Es lässt sich aber festhalten, dass in Alberta die unterschiedlichsten Kanäle zur Kopplung genutzt werden, um die Werthaftigkeit der Aufgabe Bildung innerhalb der sozialen personenbezogenen Dienstleistungsorganisation zu erzeugen und aufrecht zu erhalten.

Interessant ist zudem, dass durch die Thematisierung der Kopplungsmechanismen in der Synopse die in der best-case Studie herausgearbeiteten Faktoren eine besondere Bedeutung erfahren, die die Werthaftigkeit der Aufgaben gewährleisten und die formal lose Kopplungen anderweitig binden:

- die doppelte demokratische Legitimation des Systems,
- die strukturelle und inhaltliche Unterstützung, besonders der Kommunikationsprozesse (PLG) in den Schulen und zwischen Verwaltung und Schule
- die Professionalisierung der Akteure
- Peer Netzwerke
- das System aus „checks and balances“

So verfügt eine lange Delegationskette, die innerhalb der Provinz doppelt demokratisch abgesichert ist, über mehr Potential gesellschaftliche Ziele in die Schule zu integrieren als eine einfache demokratische Delegationskette wie sie z.B. in Deutschland vorzufinden ist.

Zudem kann die strukturelle und inhaltliche Unterstützung zwischen den Ebenen zusätzlich dafür sorgen, dass die in der Politik formulierten Ziele im Sinne der BürgerInnen interpretiert und umgesetzt werden. Professionalisierung und Peer Netzwerke koppeln die Akteure relational, bieten somit Unterstützung für Situationen, die formal lose gekoppelt erfolgen. Das System aus „checks and balances“ schließlich verhindert, dass unter dem Druck gesellschaftlicher Anpassungen Wissen der Professionen aus dem Blick gerät. Diese Praxis scheint im Einklang mit der Stewardship-Theorie zu stehen, in der davon ausgegangen wird, dass Macht durch Expertise ausgeübt wird (vgl. zusammenfassend Velte, 2010).

Da sich die beobachtete Governance im untersuchten Fall durch die intensive Zusammenarbeit und Verflechtung der Ebenen und Gremien auszeichnet, wäre hier eine Frage für zukünftigen Forschung, ob dies nicht auch Probleme mit sich bringt, wie sie in zu dichten Netzwerken beobachtet werden (hoher sozialer Druck, womöglich geringe Innovationsbereitschaft, vgl. Jansen, 2006). Die Qualität der Strukturen, der Professionalisierungsangebote und deren Offenheit für Veränderungen wären hier interessant zu untersuchen.

### *8.2 Nicht intendierte Effekte des Zentralabiturs*

In den Studien B, C und D wird gefragt, ob ex post Instrumente wie das Zentralabitur nicht intendierte, unerwünschte Effekte erzeugen. Hier wird argumentiert, dass die Einführung des Zentralabiturs in bestimmten Bereichen der Schulqualität – neben den erwünschten Effekten – auch nicht intendierte Effekte bewirkt, von denen einige zudem als unerwünscht eingeordnet werden. Als Bereiche der Schulqualität wurden die Themen Kongruenz zwischen Unterricht und Abiturprüfung, Eingrenzung der Themenvarianz (Teaching to the test), Arbeitsunzufriedenheit, individuelle und kollektive Selbstwirksamkeitserwartungen und das Schulklima untersucht.

Effekte hinsichtlich dieser Themen können erwünscht und unerwünscht sein, beherrschbar und unbeherrschbar und vorhersehbar und unvorhersehbar (Braun, 2008). Im Rahmen der Studien wurde anhand der Forschungsliteratur argumentiert, was als nicht erwünscht angesehen wird: eine Inkongruenz zwischen Unterricht und Prüfung, ein Teaching to the test, in dem Sinne, dass u.a. SchülerInneninteressen vernachlässigt werden, eine Steigerung der Arbeitsunzufriedenheit und eine sinkende Selbstwirksamkeitserwartung. Ein Effekt auf das Schulklima wurde nicht erwartet. Die formulierten Erwartungen bestätigten sich, wobei das Ausmaß der Effekte als eher gering einzuschätzen ist.

Für nicht intendierte Effekte gibt es verschiedene Ursachen und es wird hier angenommen, dass sie auch unter idealen Bedingungen auftreten, da im Bildungssystem eine interdependente Handlungssituation vorliegt und zudem Unsicherheit nicht umfassend beseitigt werden kann. An dieser Stelle wird jedoch argumentiert, dass aufgrund der vorliegenden Kopplungsmechanismen systematisch nicht intendierte, unerwünschte Effekte auftreten.

So führt das Instrument Zentralabitur einen zusätzlichen ex post Kopplungsmechanismus ein und zeitgleich eine ex ongoing Entkopplung. Es wird zu dem erst genannten angenommen, dass die Einführung eines ex post Instrumentes in eine SPDO von den Lehrenden als Vertrauensbruch interpretiert wird und einen Verdrängungseffekt zur Folge haben kann (Frey & Jegen, 2000). Die Ergebnisse zeigen, dass die Arbeitsunzufriedenheit wächst und die Wahrnehmung der individuellen und kollektiven Selbstwirksamkeit sinkt. Dies spricht z.B. dafür, zentrale Prüfungen eher als Instrument des Monitoring, also als ex ongoing Mechanismus, anstatt als ex post Mechanismus, einzuführen. Dies würde es den Lehrpersonen ermöglichen, die Ergebnisse der zentralen Prüfung, die dann während der Unterrichtszeit stattfindet, als Feedback zu nutzen. Wenn die Prüfungen darüber hinaus kein Zertifikat für die SchülerInnen umfassen, könnte dies Druck reduzieren.

Zweitens wird argumentiert, dass die eingeführte Entkopplung zwischen Lehrperson und Prüfung zu nicht intendierten Effekte führen kann, da die Lehrperson als Individuum nun unter Unsicherheit eine Interpretationsleistung erbringen muss. In diesem Sinne zeigen die Ergebnisse, dass es der Lehrperson gelingt diese Interpretation zu leisten (Inkongruenz nur sehr gering), dass aber – auch aufgrund von Unsicherheit gegenüber dem Zentralabitur – ein Teaching to the test auftritt (wie auch die weiteren Ergebnisse für das Jahr 2011 bestätigen, vgl. Oerke, Maag Merki, Maué & Jäger, 2012). Die Lehrpersonen grenzen den Unterrichtsstoff mehr als notwendig ein. Aufgrund der Ergebnisse zu Kopplungsmechanismen könnte man hier anregen, andere Arten der Kopplung als Kompensation der Entkopplung einzuführen: z.B. durch strategische, relationale oder weltanschauliche Kopplungen. In den untersuchten Bundesländern fanden z.B. Weiterbildungen statt, dies könnte positive Effekt haben. Hier konnten deren Themen, Qualität etc. leider nicht erhoben und ausgewertet werden.

Des Weiteren wurde in Studie B und C untersucht, ob Kooperation zum Curriculum einer Eingrenzung der Themenvarianz entgegenwirken kann. Dies ist interessant, da Kooperation zum Curriculum als relationaler ex ongoing Kopplungsmechanismus interpretiert werden kann, der eine Verwirklichung der Themenvarianz fördern könnte. Die Annahme wäre also, dass dieser Kopplungsmechanismus, nicht intendierte Effekte wie ein Teaching to the test minimieren könnte.

In den Studien zeigten sich divergierende Ergebnisse, die weiter überprüft werden müssen: Denn die Studie B, in der nur zentral geprüfte Kurse aus dem Jahr 2008 untersucht wurden, stellt keinen signifikanten positiven Effekt der Kooperation zum Curriculum auf die Themenvarianz fest. Im Längsschnitt zeigt sich für die zentral geprüften Kurse jedoch ein positiver Zusammenhang zwischen Kooperation zum Curriculum und Themenvarianz. Der fehlende Effekt aus Studie B spricht gegen die zuvor formulierte Annahme, dass ein relationaler, ex ongoing Kopplungsmechanismus Teaching to the test abschwächen könnte. Die Ergebnisse aus Studie C belegen dagegen die Erwartung in relationale Kopplungen. An dieser Stelle kann dies nicht weiter untersucht werden. Dies wäre eine Aufgabe für weitergehende Studien.

Einen Hinweis darauf, dass Kopplungen in relationaler Hinsicht hilfreich sein könnten, liefert das Ergebnis, dass Lehrpersonen, die von einer kollektiven Selbstwirksamkeit überzeugt sind, weniger die Themen eingrenzen als Lehrpersonen, die dies nicht annehmen (Studie C). Die Frage, die sich hier stellt, ist, wie kollektive Selbstwirksamkeitserwartungen im Kollegium unterstützt werden können. Auch hier wären weitergehende Studien interessant.

In den Bundesländern wurden auch ex ongoing Kopplungsmechanismen wie z.B. Weiterbildungen zu den zentralen Prüfungen angeboten. Die Ergebnisse aus Studie B zeigen, dass die Unterstützung (Handreichungen zu Operatoren für die Bewertung der Arbeit, Themenvorgaben und Fortbildungen) zumindest im Nachhinein – von den Lehrpersonen für die Herstellung der Kongruenz zwischen Vorbereitung im Unterricht und der Abiturprüfung positiv in einem Zusammenhang stehen (Studie C). Hinsichtlich der Eingrenzung der Themenvarianz wurde dies aber nicht untersucht. Studien zu diesen und anderen Kopplungsmechanismen als Ausgleich wären hier spannend. Zudem liegen bisher z.B. keine Aussagen zur Qualität und Quantität der Weiterbildung vor. Auch dies



könnte weitergehend untersucht werden, um zu erfahren, inwiefern relationale Kopplungsmechanismen nicht intendierte Effekte von ex post Kopplungsmechanismen abfedern könnten.

## **9 Fazit und Ausblick**

In der Synopse wurde erstens gefragt, welche Arten der Kopplungsmechanismen dazu beitragen können, dass in Bildungssystemen die erwünschte Leistung erbracht wird. Die Fokussierung der Kopplungsmechanismen in der best-case Studie zeigt, dass Evaluationen und Inspektionen vor allem ex ongoing eingesetzt werden und dass ex ante Mechanismen im best-case eine wichtige Rolle einnehmen. Im untersuchten Fall können die ex ongoing Mechanismen als Teil einer konstruktiven Fehlerkultur interpretiert werden. Dies bedeutet, dass nicht intendierte Effekte kontinuierlich aufgedeckt werden und versucht wird, ihnen durch Koordination entgegen zu wirken. Statt auf Sanktionen wird auf Unterstützung gesetzt. In einem solchen System, so könnte interpretiert werden, gehört das gegenseitige Vertrauen wie die Annahme, dass Fehler, also unerwünschte Effekte, zur Steuerung dazugehören zum Bestandteil der Arbeit von Lehrpersonen, Schulleitung und Verwaltung. Dies müsste jedoch weitergehend untersucht werden. Das Fallbeispiel zeigt schließlich, dass ex post Instrumente die Möglichkeit bieten, Handlungen zu sanktionieren, die sozusagen die Toleranzfehlerschwelle überschreiten. Im untersuchten best-case werden demnach Kopplungsmechanismen kombiniert, zudem bieten Professionelle Lerngemeinschaften die Möglichkeit, dass u.a. Lehrpersonen und Schulleitungen etc. multiplex mit anderen Akteuren gekoppelt sind.

Zweitens wurde in der Synopse gefragt, wie sich Reformen der Kopplungsmechanismen auswirken. Die Studien zur Implementation des Zentralabiturs untersuchten nicht intendierte, unerwünschte Effekte. Das Ergebnis ist, dass unerwünschte Effekte auftreten und im Zusammenhang mit der Einführung zentraler Prüfungen stehen. Die Erörterungen zu den Kopplungsmechanismen regen dazu an, nicht intendierte, unerwünschte Effekte durch zusätzliche Kopplungsmechanismen abzufedern. Auch bietet der Ansatz der Kopplungsmechanismen Möglichkeiten, über Alternativen eines Instrumentes nachzudenken. Denn Instrumente und ihre Wirkung können leicht variiert werden, wenn sie beispielsweise auf der zeitlichen Dimension neu justiert werden (z.B. wenn zentrale Prüfungen nicht ex post stattfinden, sondern ex ongoing).

Die vorliegende Synopse hat zudem Typologien von Kopplungsmechanismen referiert und diskutiert. Hier besteht weiterhin ein großer Forschungsbedarf. Interessant um die Typologien weiterzuentwickeln erscheint m.E. die Netzwerktheorie (vgl. u.a. Jansen, 2006) und die Theorie der Relativität des Sozialen (Langer, 2009a). Aufgrund letzterer wurde bereits eine Methodik der Mechanismen-Analyse vorgestellt (Langer, 2009b). Dies scheinen interessante weitergehende Ansätze für die Forschung zu sein.

Weiterer Forschungsbedarf besteht zur Qualität von Kopplungen und Kopplungsmechanismen. In den Studien zum Zentralabitur wurden z.B. keine Weiterbildungsangebote analysiert, obwohl die Qualität und Ausrichtung dieser sicher-

lich einen wichtigen Beitrag zum Gelingen des Instrumentes Zentralabitur leistet. In der best-case Analyse konnte auch nicht die Qualität der einzelnen Netzwerke untersucht werden.

Zudem wären weitere Fallanalysen spannend. Im Rahmen der genannten Studie A steht z.B. die Auswertung von drei weiteren Fallstudien in anderen Provinzen noch aus. Ein Vergleich der Kopplungsmechanismen wäre hier sicherlich interessant. Zudem wären Arbeiten, die die aufgeworfenen Hypothesen z.B. zur doppelten demokratischen Legitimation quantitativ überprüfen, zu begrüßen.

Methodisch wären in den Zentralabiturstudien hinsichtlich der untersuchten Faktoren z.B. noch fachspezifische Effekte von Teaching to the test interessant. Auch wären Pfadmodellanalysen spannend, die die Rolle der Unsicherheit gegenüber den zentralen Prüfungen näher in den Blick nehmen. Außerdem stellt sich hinsichtlich des Zentralabiturs die Frage, ob und unter welchen Bedingungen die positiven Effekte des Zentralabiturs die hier aufgezeigten unerwünschten Effekte überwiegen. Dies konnte im Rahmen dieser Arbeit nicht diskutiert werden. Schliesslich wäre es inhaltlich spannend, mehr über Fehlerkulturen an Schulen und in Bildungssystemen zu erfahren.

## Literaturverzeichnis

- Alberta Education. (2006). *Professional learning communities: an exploration*, Alberta, Canada. Verfügbar unter <http://www.msdc-mn.org/files-documents/ProfLearnExplore.pdf>.
- Altrichter, H. & Heinrich, M. (2007). Kategorien der Governance-Analyse und Transformationen der Systemsteuerung in Österreich. In H. Altrichter, T. Brüsemeister & J. Wissinger (Hrsg.), *Educational Governance. Handlungskoordination und Steuerung im Bildungssystem* (S. 55–103). Wiesbaden: Verlag für Sozialwissenschaften | GWV Fachverlage GmbH Wiesbaden.
- Altrichter, H. & Helm, C. (2011). Schulentwicklung und Systemreform. In H. Altrichter & C. Helm (Hrsg.), *Akteure & Instrumente der Schulentwicklung* (Professionswissen für Lehrerinnen und Lehrer, Bd. 7, S. 13–37). Zürich: Pestalozzianum.
- Altrichter, H. & Maag Merki, K. (2010). Steuerung der Entwicklung des Schulwesens. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch neue Steuerung im Schulsystem* (Educational governance, Bd. 7, S. 15–39). Wiesbaden: Verlag für Sozialwissenschaften.
- Andrews, S. (2004). Washback and curriculum innovation. In L. Cheng and Y. Watanabe (Hrsg.), *Washback in language testing: Research contexts and methods*, S. 37–50. Mahwah, NJ: Lawrence Erlbaum.
- Au, W. (2007). High stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher* 36 (5): 258–67.
- Bellmann, J. & Weiß, M. (2009). Risiken und Nebenwirkungen Neuer Steuerung im Schulsystem. Theoretische Konzeptualisierung und Erklärungsmodelle. *Zeitschrift für Pädagogik*, 55 (2), 286–308. Verfügbar unter <http://www.pedocs.de/volltexte/2011/4251>; <http://nbn-resolving.de/urn:nbn:de:0111-opus-42518>.
- Berkemeyer, N. & Manitz, V. (2013). Gerechtigkeit als Kategorie der Analyse von Schulsystemen - das Beispiel Chancenspiegel. In K. Schwippert, M. Bonsen & N. Berkemeyer (Hrsg.), *Schul- und Bildungsforschung. Diskussionen, Befunde und Perspektiven. Festschrift für Wilfried Bos* (S. 223–241). Münster: Waxmann.
- Berkemeyer, N. (2010). *Die Steuerung des Schulsystems. Theoretische und praktische Explorationen*. 1. Aufl. Wiesbaden: VS, Verlag für Sozialwissenschaften.
- Berryhill, J., Linney, J.A. & Fromewick, J. (2009). The effects of education accountability on teachers: are policies too-stress provoking for their own good? *International Journal of Education Policy and Leadership*, 4 (5), 1-14.
- Bishop, J. H. (1999). Are national exit examinations important for educational efficiency? *Swedish Economic Policy Review* (6), 349–398.
- Blömeke, S., Herzig, B. & Tulodziecki, G. (2007). *Gestaltung von Schule. Eine Einführung in Schultheorie und Schulentwicklung*. Bad Heilbrunn: Klinkhardt, Julius.
- Blömeke, S. & H. B. (2009). Schule als gestaltete und zu gestaltende Institution – ein systematischer Überblick über aktuelle und historische Schultheorien. In S. Blömeke, T. Bohl, L. Haag, G. Lang-Wojtasik & W. Sacher (Hrsg.), *Handbuch Schule*.

- Theorie - Organisation - Entwicklung* (UTB, Bd. 8392, S. 15–28). Bad Heilbrunn: Klinkhardt.
- Bonsen, M. & Rolff, H.-G. (2006). Professionelle Lerngemeinschaften von Lehrerinnen und Lehrern. *Zeitschrift für Pädagogik*, 52 (2), 167–184.
- Bös, M. (2002). *Zu den Grenzen der politischen Steuerung von Grenzen. Nicht-intendierte Folgen von Einwanderungs- und Staatsangehörigkeitsgesetzgebung*. Zugriff am 04.05.2013. Verfügbar unter [http://www.uni-marburg.de/fb03/soziologie/institut/arbeitschwerpunkte/ang\\_soz/boes/publikationenorder/grenzen\\_politischer\\_steuerung.pdf](http://www.uni-marburg.de/fb03/soziologie/institut/arbeitschwerpunkte/ang_soz/boes/publikationenorder/grenzen_politischer_steuerung.pdf).
- Bos, W. & Schwippert, K. (2002). TIMSS, PISA, IGLU & Co. Vom Sinn und Unsinn internationaler Schulleistungsuntersuchungen. *Bildung und Erziehung*, 55, 5–23.
- Braun, D. (2008). Evaluation und unintendierte Effekte - eine theoretische Reflexion. In H. Matthies & D. Simon (Hrsg.), *Wissenschaft unter Beobachtung. Effekte und defekte von Evaluationen* (Leviathan Sonderhefte, S. 103–124). VS Verlag für Sozialwissenschaften.
- Brockmeyer, R. & Edelstein, W. (1997). *Selbstwirksame Schulen: Wege pädagogischer Innovation*. Oberhausen: Laufen.
- DIPF (Hrsg.). (2007). *Schulleistungen und Steuerung des Schulsystems im Bundesstaat: Kanada und Deutschland im Vergleich*. Frankfurt am Main
- Döhler, M. (2005). Die begrenzte Rationalität von Delegation und Steuerung in der Bundesverwaltung. In S. Ganghof & P. Manow (Hrsg.), *Mechanismen der Politik. Strategische Interaktion im deutschen Regierungssystem* (Schriftenreihe des Max-Planck-Instituts für Gesellschaftsforschung Köln, Bd. 54, S. 215–245). Frankfurt/Main: Campus-Verlag
- Donaldson, L. (1990). The ethereal hand: Organizational economics and management theory. *Academy of Management Review*, 15 (3), 369–381.
- Drepper, T. & Tacke, V. (2012). Die Schule als Organisation. In M. Apelt & V. Tacke (Hrsg.), *Handbuch Organisationstypen* (S. 205–237). Wiesbaden: Springer Fachmedien. Zugriff am 04.05.2013. Zugriff am 04.05.2013.
- Drepper, T. (2010). Soziale personenbezogene Dienstleistungsorganisationen aus neoinstitutionalistischer Perspektive. In T. Klatetzki (Hrsg.), *Soziale personenbezogene Dienstleistungsorganisationen. Soziologische Perspektiven* (Organisation und Gesellschaft, 1. Aufl., S. 129–165). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Evans, J. E. (1997). *Organizational and individual responses to educational reforms in Alberta* (Alberta University, Hrsg.) (Reports), Edmonton
- Fend, H. (1988). Schulqualität. Die Wiederentdeckung der Schule als pädagogische Gestaltungsebene. *Neue Sammlung*, 28, 537–547. Zugriff am 16.06.2013.
- Fend, H. (1998). *Qualität im Bildungswesen: Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. Weinheim: Juventa.
- Fend, H. (2008). *Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität*. Wiesbaden: Verlag für Sozialwissenschaften.

- Fend, H. (2009). *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen*. Wiesbaden: Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden.
- Flick, U. (2008). *Triangulation. Eine Einführung* (Qualitative Sozialforschung, Bd. 12, 2. Aufl.). Wiesbaden: Verlag für Sozialwissenschaften.
- Frey, B. S. & Jegen, R. (2002). Kontraproduktive Wirkung des Motivators "Geld". Intrinsische Motivation. *Verbands-Management*, 28 (3), 30–41.
- Gilardi, F. & Braun, D. (2002). Delegation aus der Sicht der Prinzipal-Agent-Theorie. *Politische Vierteljahresschrift*, 43 (1), 147–161.
- Glassman, R.B. (1973): Persistence and loose coupling in living systems. In: *Behavioral Science*, 18, S. 83 – 98.
- Hox, J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah NJ [u.a.]: Erlbaum.
- Jäger, D. J. (2012a). Herausforderung Zentralabitur: Unterrichtsinhalte variieren und an Prüfungsthemen anpassen. In K. Maag Merki (Hrsg.), *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland*. (S. 175–203). Wiesbaden: Springer.
- Jäger, D. J. (2012b). Schulklima, Selbstwirksamkeit und Arbeitszufriedenheit aus Sicht der Lehrpersonen und Schüler/-innen in Hessen und Bremen. In K. Maag Merki (Hrsg.), *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland*. (S. 61–91). Wiesbaden: Springer.
- Jäger, D. J. (2014). Zwischen Empowerment und Kontrolle. Die praktische Umsetzung des New Public Management und Professioneller Lerngemeinschaften in Kanada. Eine Fallstudie. *Zeitschrift für Pädagogik*, 60 (2), 303–324
- Jäger, D. J., Maag Merki, K., Oerke, B. & Holmeier, M. (2012). Statewide low-stakes tests and a teaching to the test effect? An analysis of teacher survey data from two German states. *Assessment in Education: Principles, Policy & Practice*, 19 (4), 451–467.
- Jansen, D. (2006). *Einführung in die Netzwerkanalyse. Grundlagen, Methoden, Forschungsbeispiele* (Lehrbuch, 3. Aufl.). Wiesbaden: Verlag für Sozialwissenschaften.
- Klafki, W. (2002). *Schultheorie, Schulforschung und Schulentwicklung im politisch-gesellschaftlichen Kontext. Ausgewählte Studien*. Weinheim, Basel: Beltz.
- Kühn, S. M. (2012). Zentrale Abiturprüfungen im nationalen und internationalen Vergleich mit besonderer Perspektive auf Bremen und Hessen. In K. Maag Merki (Hrsg.), *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland* (S. 25–43). Wiesbaden: Springer.
- Langer, R. & Hüther, O. (2009). Das Governance-Konzept. In R. von Lüde, D. Moldt & R. Valk (Hrsg.), *Selbstorganisation und Governance in künstlichen und sozialen Systemen* (Wirtschaft - Arbeit - Technik, Bd. 5, S. 469–486). Münster: LIT Verlag.
- Langer, R. (2009a). Die Relativität des Sozialen. In R. von Lüde, D. Moldt & R. Valk (Hrsg.), *Selbstorganisation und Governance in künstlichen und sozialen Systemen* (Wirtschaft - Arbeit - Technik, Bd. 5, S. 119–153). Münster: LIT Verlag.
- Langer, R. (2009b). Methodologie der Mechanismen-Analyse. In R. von Lüde, D. Moldt & R. Valk (Hrsg.), *Selbstorganisation und Governance in künstlichen und so-*

- zialen Systemen (Wirtschaft - Arbeit - Technik, Bd. 5, S. 153–192). Münster: LIT Verlag.
- Langer, R. (2009c). Governance als Analyseinstrument. In R. von Lüde, D. Moldt & R. Valk (Hrsg.), *Selbstorganisation und Governance in künstlichen und sozialen Systemen* (Wirtschaft - Arbeit - Technik, Bd. 5, S. 487–536). Münster: LIT Verlag.
- Lasky, S., Datnow, A. & Stringfield, S. (2005). Linkages between federal, state and local levels in educational reform. In N. Bascia, A. Cumming, A. Datnow, K. Leithwood & D. Livingstone (Hrsg.), *International handbook of educational policy* (Springer international handbooks of education, Bd. 13, S. 239–259). Dordrecht: Springer.
- Maag Merki, K. & Oerke, B. (2012). Methodische Grundlagen der Studie. In K. Maag Merki (Hrsg.), *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland* (S. 43–61). Wiesbaden: Springer.
- Manow, P. (2005). Die politische Kontrolle der Ministerialbürokratie des Bundes: Die Bedeutung der Landesebene. In S. Ganghof & P. Manow (Hrsg.), *Mechanismen der Politik. Strategische Interaktion im deutschen Regierungssystem* (Schriftenreihe des Max-Planck-Instituts für Gesellschaftsforschung Köln, Bd. 54, S. 245–277). Frankfurt/Main: Campus-Verlag.
- March, J. G. & Olsen, J. P. (1984). The new institutionalism: organizational factors in political life. *The American Political Science Review*, 78 (3), 734–749.
- Mayntz, R. (2002). *Akteure - Mechanismen - Modelle. Zur Theoriefähigkeit makrosozialer Analysen* (Schriften des Max-Planck-Instituts für Gesellschaftsforschung Köln 42). Frankfurt: Campus-Verlag.
- Mayntz, R. (2004). *Governance Theory als fortentwickelte Steuerungstheorie?*: MPIfG Working Paper. Zugriff am 12.04.2013. Verfügbar unter <http://www.mpifg.de/pu/workpap/wp04-1/wp04-1.html>.
- Mayntz, R. (2008). Von der Steuerungstheorie zu Global Governance. In G. F. Schuppert & M. Zürn (Hrsg.), *Governance in einer sich wandelnden Welt* (Politische Vierteljahresschrift, Bd. 41). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Merton, R. K. (1936). The unanticipated consequences of purposive social action. *American Sociological Review*, 6 (1), 894–904. Zugriff am 05.05.2013.
- Meyer, J. W. & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83 (2), 340–363.
- OECD (2004). *Bildung auf einen Blick. OECD-Indikatoren 2004*. Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- OECD (2007). PISA 2006. Science Competencies for Tomorrow's World. Paris: OECD.
- OECD (2010). Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften (PISA 2009 Ergebnisse, Bd. 1). Bielefeld: Bertelsmann.
- Oerke, B., Maag Merki, K., Maué, E. & Jäger, D. J. (2012). Zentralabitur und Themenvarianz. Lohnt sich Teaching-to-the-Test? In D. Bosse, F. Eberle & B. Schneider-Taylor (Hrsg.), *Standardisierung in der gymnasialen Oberstufe* (S. 27–49). Dordrecht: Springer.

- Offe, C. (2008). Empty Signifier oder sozialwissenschaftliches Forschungsprogramm? In G. F. Schuppert & M. Zürn (Hrsg.), *Governance in einer sich wandelnden Welt* (Politische Vierteljahresschrift, Bd. 41, S. 61–76). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Osborne, M. J. (2004). *An introduction to game theory*. Oxford: Oxford University Press.
- Portes, A. (2000). The hidden abode: sociology as analysis of the unexpected. *American Sociological Review*, 65 (1), 1–18.
- Raudenbush, S. W., Bryk, A. S. & Congdon, R. (2004). HLM 6 for Windows [Computer software]. Lincolnwood Ill: Scientific Software International.
- Richter, P. (2012). Die Organisation öffentlicher Verwaltung. In M. Apelt & V. Tacke (Hrsg.), *Handbuch Organisationstypen* (S. 91–113). Wiesbaden: Springer Fachmedien.
- Rolff, H.-G. (2001). Professionelle Lerngemeinschaften. Eine wirkungsvolle Synthese von Unterrichts- und Personalentwicklung. In H. Buchen & L. Horster (Hrsg.), *Schulleitung und Schulentwicklung Erfahrungen, Konzepte, Strategien* (D 6.5). Berlin: Raabe Fachverlag für Bildungsmanagement.
- Rolff, H.-G. (2007). *Studien zu einer Theorie der Schulentwicklung*. Weinheim [u.a.]: Beltz.
- Sauermann, J. & Kaiser, A. (2010). Taking others into account: SelfInterest and fairness in majority decision making. *American Journal of Political Science*, 54 (3), 667–685. Zugriff am 21.08.2013.
- Scharpf, F. W. (2000). *Interaktionsformen. Akteurzentrierter Institutionalismus in der Politikforschung* (UTB 2136 : Politikwissenschaft). Opladen: Leske + Budrich.
- Schedler, K. & Proeller, I. (2006). *New Public Management* (UTB 2132 : Public Management Betriebswirtschaft, 3. Aufl.). Bern: Haupt.
- Schimank, U. (2007). Die Governance-Perspektive: Analytisches Potenzial und anstehende konzeptionelle Fragen. In H. Altrichter, T. Brüsemeister & J. Wissinger (Hrsg.), *Educational Governance. Handlungskoordination und Steuerung im Bildungssystem* (S. 231–260). Wiesbaden: Verlag für Sozialwissenschaften | GWV Fachverlage GmbH Wiesbaden.
- Schmitz, G.S. & Schwarzer, R. (2000). Selbstwirksamkeitserwartung von Lehrern: Längsschnittbefunde mit einem neuen Instrument. *Zeitschrift für Pädagogische Psychologie*, 14 (1), 12–25.
- Schuppert, G. F. (2008). Governance - auf der Suche nach Konturen eines "anerkannt uneindeutigen Begriffs". In G. F. Schuppert & M. Zürn (Hrsg.), *Governance in einer sich wandelnden Welt* (Politische Vierteljahresschrift, Bd. 41, S. 13–40). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Strøm, K. (2000). Delegation and accountability in parliamentary democracies. *European Journal of Political Research*, 37 (3), 261–289.
- Tarnoczi, J. (2006). Critical reflections on professional learning communities in Alberta. *Electronic Journal of Sociology*. Verfügbar unter: <http://www.sociology.org/content/2006/tier2/tarnoczi.html>; Stand: 10.12.12

- Thillmann, K. (2011). Reorganisation von Schulen unter den Bedingungen der Neuen Steuerung. In F. Dietrich, M. Heinrich & N. Thieme (Hrsg.), *Neue Steuerung - alte Ungleichheiten? Steuerung und Entwicklung im Bildungssystem* (S. 65–77). Münster: Waxmann.
- Tillmann, K.-J., Dederich, K., Kneuper, D., Kuhlmann, C. & Nessel, I. (2008). *PISA als bildungspolitisches Ereignis*: VS Verlag für Sozialwissenschaften (GWV).
- Van Ackeren, I. & Brauckmann, S. (2010). Internationale Diskussions-, Forschungs- und Theorieansätze zur Governance im Schulwesen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch neue Steuerung im Schulsystem* (Educational governance, Bd. 7, S. 41–61). Wiesbaden: VS Verl. für Sozialwissenschaften.
- Velte, P. (2010). Stewardship-Theorie. *Zeitschrift für Planung & Unternehmenssteuerung*, 20 (3), 285–293.
- Wald, A. & Jansen, D. (2007). Netzwerke. In A. Benz, S. Lütz & G. Simonis (Hrsg.), *Handbuch Governance. Theoretische Grundlagen und empirische Anwendungsfelder* (S. 93–106). Wiesbaden: Verlag für Sozialwissenschaften.
- Wehling, P. (2008). Wissen und seine Schattenseite: Die wachsende Bedeutung des Nichtwissens in (vermeintlichen) Wissensgesellschaften. In T. Brüsemeister & K.-D. Eubel (Hrsg.), *Evaluationsbasierte Steuerung, Wissen und Nichtwissen* (S. 17–34). Wiesbaden: Verlag für Sozialwissenschaften.
- Weick, K. E. (1976/ 2009). Bildungsorganisationen als lose gekoppelte Systeme. (Übersetzt und erneut aufgelegt: Originaltitel: „Educational Organizations as Loosely Coupled Systems“ von Karl E. Weick, publiziert in der „Administrative Science Quarterly“ (1976), 21, 1-19 mit Erlaubnis von „Administrative Science Quarterly“, Volume 21; Johnson Graduate School of Management, Cornell University. Übersetzung: Michael Schemmann.) In S. Koch & M. Schemman (Hrsg.), *Neo-Institutionalismus in der Erziehungswissenschaft. Grundlegende Texte und empirische Studien* (S. 85–109). Wiesbaden: Verlag für Sozialwissenschaften.



Daniela J. Jäger (2014): Zwischen Empowerment und Kontrolle. Die praktische Umsetzung des New Public Management und Professioneller Lerngemeinschaften in Kanada. Eine Fallstudie. *Zeitschrift für Pädagogik*, 60 (2), 303-324.

Hinweis

Sperrfrist bis 01.05.2015.

Daniela J. Jäger, Katharina Maag Merki, Britta Oerke & Monika Holmeier (2012): Statewide low-stakes tests and a teaching to the test effect? An analysis of teacher survey data from two German states. *Assessment in Education: Principles, Policy & Practice*, 19 (4), 451-467.

#### Hinweis

This is an Author's Accepted Manuscript of an article published in *Assessment in Education: Principles, Policy & Practice*, Volume 19, Issue 4, 2012, Published online: 20. April 2012, © 2012 Taylor & Francis, available at <http://www.tandfonline.com/10.1080/0969594X.2012.677803>.

# **Statewide Low-Stakes Tests and a Teaching-to-the-Test Effect? An Analysis of Teacher Survey Data from Two German States**

Daniela J. Jäger, Katharina Maag Merki, Britta Oerke, Monika Holmeier

*Institute of Education, University of Zurich, Zurich, Switzerland*

Email: djaeger@ife.uzh.ch  
University of Zurich  
Institute of Education  
Freiestrasse 36  
8032 Zurich  
Switzerland

Governments worldwide are changing their educational policies to improve student performance. Some of these reforms involve the introduction of high-stakes testing, which often creates negative effects such as “teaching to the test.” This article deals with the use of low-stakes testing in the German states of Hesse and Bremen, and analyses whether statewide low-stakes testing produces comparable results to the use of high-stakes tests from the perspective of teachers. Moreover, it seeks to identify the factors responsible for the observable teaching-to-the-test effect, above and beyond the use of statewide exams. The data come from surveys carried out in the framework of a longitudinal study funded by the German Research Foundation (DFG). Hesse and Bremen introduced statewide exit exams in several subjects in 2007 and 2008 as low-stakes tests. Our quasi-experimental study has been evaluating these implementations since 2007.

Keywords: Statewide exams; Teaching-to-the test; Germany

## **Introduction**

Defining statewide tests is difficult because they vary in many ways. They are administered in different types of schools at different stages, their content as well as the exam procedures differ, and the results may or may not affect students, schools, and teachers (high-stakes or low-stakes exams [Madaus 1988: 87, cited in Andrews 2004]). In an analysis of 15 OECD countries<sup>1</sup>, Klein, Kühn, van Ackeren, and Block (2009: 602) concluded that statewide exams have only one thing in common: that all students are assigned the same test questions at the same time.

It is therefore no surprise that findings on the intended and unintended effects of statewide tests differ. One of these unintended effects is the teaching-to-the-test effect. Generally, this phenomenon is defined as a narrowing of teaching content to an

undesirable extent and is reported primarily in connection with high-stakes, statewide exit exams (Cheng & Curtis 2004: 3). It has been suggested that this unintended effect of high-stakes tests can be avoided with low-stakes tests, which allow the continued use of output management. Abrams and Madaus (2003) and Brozo and Hargis (2003) hypothesise that low-stakes exit exam testing might increase instructional quality. Theoretically, their argument is plausible: Since high-stakes tests entail consequences for teachers and/or schools, they generate pressure that leads teachers to “overconform”, narrowing the curriculum more than is desirable. On this subject, Hamilton and Stecher (2002: 91), making reference to Corbett and Wilston (1991), propose the hypothesis that “the greater the stake, the more likely that [...] narrowing would occur.”

Germany is one of the 15 OECD countries in which the standardisation of statewide exams can be considered low to moderate, as shown by Klein et al. (2009: 615). In the German context, low to moderate standardization means that statewide exams are administered in just three to six subjects, that teachers or students are allowed to select the test questions, that grading is generally done by a teacher from the same school, and in most cases, that the exams are not graded on an anonymous basis. In Hesse and Bremen, teachers are given grading guidelines. But these vary again in their details and depending on the subject. Finally, the exam results are published only to a limited extent (e.g., in Hesse<sup>2</sup> only for the state as a whole) or not at all (e.g., in Bremen). Moreover, statewide exams in Germany are low-stakes for schools and teachers because neither of the two are threatened with sanctions. They are also low-stakes for students because the exam grades are based on the outcomes of the statewide exit exam and on student achievement in the last two years of school. In Hesse, the exit exam makes up just 16% and in Bremen just 23% of the overall final

grade (Klein et al. 2009: 608). Of course, one can argue that the overall result is high-stakes for the students, as grades are the most important selection criterion used by universities. But in comparison to other countries, the statewide exit exams in Germany are characterised in general by low pressure on students (just 16% until 23% of the overall final grade) and consequently low exam stakes (Klein et al. 2009).

Although only low-stakes exams were introduced in Hesse and Bremen, these exams raised the stakes overall relative to the prior situation, which may have an impact on teacher behaviour. The current article presents the results of a study conducted in the two German states of Hesse and Bremen. It addresses two questions:

1. Do the statewide low-stakes exit exams produce the effect of teaching to the test from the teachers' point of view?
2. What factors, aside from the use of statewide exit exams, are responsible for teaching to the test?

In the following, we present our theoretical framework and the results of our empirical study. Our data set consists of teacher and student surveys carried out in a longitudinal study. In the year 2007, Hesse and Bremen introduced statewide high-school exit exams in various subjects. Since Bremen is introducing the statewide exit exams incrementally and not in all subjects, it has been possible to compare teachers' responses in classes with class-based exams and those with statewide exams. In this article, we present descriptive findings from this study from 2007 and 2008 and conduct multilevel analyses.

### **Variations of Teaching to the Test (TTT)**

Influences of testing on teaching and learning are described as “washback” or “backwash” effects in the literature (Alderson & Wall 1993: 117; cf. Cheng & Watanabe 2004). These terms have a neutral connotation, similar to “influence” (ibid.). “Teaching to the test” as a washback effect can be observed in various forms.

According to Abrams (2007: 90), “Classroom assessments mirror the format, item types, and, in some cases, the cognitive demands of the state test.” In a meta-analysis Au (2007: 262) presented evidence of three types of TTT. In the case of “content control”, lesson content is aligned to the material that will be covered in the test; “formal control” results in a fragmentation of knowledge in which content is broken down and taught in small, separate pieces; “pedagogic control” entails an increase in teacher-centred instruction. Additionally, curriculum reallocation strategies have been examined in the context of specific subjects: Hamilton & Stecher (2002: 92-93), for instance, showed that “the size of subject-to-subject shifts in emphasis can be substantial.”

Other studies have identified effects on teachers’ behaviour toward students in grading exams. Clarke, Haney, and Madaus (2000) reported evidence from US states as well as Ireland indicating that high-stakes testing programs are linked to increased dropout and grade retention rates and decreased rates of completion (see also the Position Statement of the American Evaluation Association 2009). Analogously to the TTT effect, this reaction to statewide exams could be called “grading to the test” and aims to prevent students with low achievement from participating in statewide exams.

On the one hand, a certain degree of focus in exams is desirable, for instance, to harmonise teachers’ behaviour (Hoppe 2004, Richter 2004). On the other hand, however, “grading to the test” and the teaching of exam material without conveying the meaning or relevance of the subject or topic are seen as negative effects (Clarke et al. 2000; American Evaluation Association 2009, Valli, Croninger, Chambliss, Graeber & Buese 2008).

In this article, “teaching to the test” is considered an unintended or negative effect of statewide testing, whereby the range of curricular content is narrowed to

such an extent that students' needs are overlooked and teachers' freedom to tailor their curriculum is severely abridged.

### **Theoretical Framework**

New control mechanisms currently being introduced in educational systems worldwide—particularly decentralization and recentralization measures—aim at improving student performance (see OECD 2004). These new public management (NPM) reforms establish output indicators for the purpose of measuring performance and ensuring accountability. Behaviour that conforms to the requirements formulated in the output indicators is rewarded, whereas non-conforming behaviour is sanctioned. This “feedback loop” (Hamilton, Stecher, Russell, Marsh & Miles 2008) based on rewards and incentives is designed to ensure that the “agent” conducts the tasks assigned by the “principal” (principal-agent theory, see Niskanen 1971).

Hoppe (2004) and Richter (2004) put forward an argument in favour of these reforms: The instrument of statewide exams provides a basis for ex-post evaluation of teacher performance. Furthermore, Bishop (1995: 658) argues that high-stakes tests increase the motivation of students, parents, educational administrators, and teachers by introducing peer pressure (Bishop 1995: 658) and by increasing extrinsic motivation. From this point of view, focusing the curriculum on the statewide exam is, to a certain extent, desirable.

Here, three counterarguments should be considered: First, students' underperformance in achievement tests is not due solely to a lack of motivation (some of the most important factors include social background, cultural practice, and joy of reading; OECD, 2010). Second, neither incentives nor external pressure can guarantee that the policy goal of “lifelong learning” will be achieved (EU Commission 2009). Individual initiative is crucially required as well. Statewide high-stakes exit exams

are, however, understood as an external incentive. The third argument, which can be considered the main detrimental effect of limiting teachers' freedom in curriculum design, is that achievement tests undermine self-regulation (Deci, Koestner & Ryan 1999: 633). With our third counterargument to the assumption that teachers and school administrators need incentives to do their work well, we refer to Ulich (1994), who argues that discretionary leeway for employees has an impact on performance improvement. That is, that teacher discretion improves performance and that teachers should therefore be granted a certain degree of freedom in choosing lesson topics. Empirical evidence supporting this hypothesis has been presented in several studies; see, for instance, Jordan (1986), Ardan, Quarstein, and McAfee (1994), and Glaser, Herbig, and Gunkel (2008). Frey and Jegen explain these findings in their summary of the "motivation crowding theory" (2000: 6-7), which posits two opposing effects of external intervention on the agent's performance: a "crowding-in" effect, whereby external intervention raises performance by increasing the marginal monetary benefit of performing better, and a "crowding-out" effect, whereby "external intervention undermines intrinsic motivation and thus negatively affects the agent's marginal benefit from performing. Stronger external intervention unequivocally reduces the agent's performance level, provided the disciplining effect does not work." Frey and Jegen (2000: 7) conclude that, from the principal's point of view, the beneficial effect of intervention depends on the relative magnitudes of these two countervailing effects. Moreover, Hacker (2005) argues that the perception of freedom should be divided into objective, formally given degrees of freedom and subjective, perceived degrees of freedom. In reality, of course, the objective degrees of freedom do not necessarily coincide with the subjective ones. That is, teachers' perceptions determine whether they use their discretion or not.



We therefore conclude that external, top-down interventions, which teachers may perceive just as added control and restriction of their freedom, may have negative consequences on both teaching and learning.

### **Statewide Exams in Hesse and Bremen**

In Germany, many schools introduced statewide exams and the concept of school autonomy simultaneously to ensure that important content would not be neglected in the process of granting schools greater freedom. The statewide exams were therefore implemented primarily in the form of “content standards” (Klieme, Avenarius, Blum, Döbrich, Gruber & Prenzel 2003): They aim to test a specific, predefined content and to guarantee the comparability of exam results (KMK 2005: 9). In Hesse and Bremen, these content standards were implemented by giving teachers descriptions of exam topics about two years in advance of the exam, with the descriptions varying in the level of detail from one subject to the next. Apart from the information about the exam topics, teachers were provided with mandatory syllabi including exam topics and educational goals that elaborate on the teaching guidelines.<sup>3</sup> As this example shows, the intention in these two states was not to restrict the curriculum to the topics covered on the statewide exam. It should also be noted that the state has not created either incentives or sanctions linked to the statewide exam results. This shows that teachers are formally permitted to exercise a certain degree of professional discretion in curriculum design.

### **Do Statewide Low-Stakes Exit Exams Produce a Teaching-to-the-Test Effect?**

To answer the first research question as to whether a teaching-to-the-test effect or a “washback” effect has occurred in Hesse and Bremen, we refer to five context factors cited by Watanabe (2004: 22): macro context, micro context, personal factors,

prestige factors, and test factors. In the literature, eleven instruments are mentioned that specify these factors:

**\*\*\*Table 1. Instruments that may affect the TTT effect\*\*\***

**\*\*\* about here \*\*\***

Our first hypothesis concerning the first question (Q1H1) is that we expect a slight TTT effect. As the statewide exams in Hesse and Bremen have no consequences for the teacher, the school, or the school district, and as neither external examiners nor anonymous grading are used, we would not expect to a TTT effect. However, as the reform factually raises the stakes by implementing statewide exams and as the exam type can be described most accurately as “content standards,” a TTT effect may appear.

Hypotheses Q1H2 and Q1H3 imply a distinction between current participation in statewide exams and prior experience with statewide exams. We assume, on the one hand, that teachers who are currently administering statewide exams narrow the curriculum to a higher degree than teachers who are not (Q1H2). On the other hand, with respect to the question of teachers’ past experience with statewide exams, we assume that teachers are unable to profit from the learning effect of prior experience, as new exam topics are selected each year (Q1H3). As a result, we expect no difference between teachers with one and two years of experience. At the school level, we consider the state to be a confounding variable. Here, no hypothesis is formulated.

**What Factors are Responsible for Teaching to the Test?**

To analyse the factors responsible for teaching to the test, we study uncertainty regarding statewide exams, teachers' experience with the statewide testing system, teacher cooperation on the curriculum, and individual self-efficacy, all at the individual level. We also expect school-level aspects to have an effect (state, student abilities).

We assume that uncertainty plays a decisive role in the perception of narrowed curricular content (Q2H1). As uncertainty is an important factor that can affect perceptions of teaching (Liermann 2009: 246), we assume that uncertainty concerning the requirements of the statewide exams is an innate characteristic of the statewide exam as a testing instrument. Theoretically, statewide exams are designed to guarantee objectivity by introducing an outside and, it is hoped, more objective person or institution for task design and evaluation. At the same time, this results in an information gap because the person teaching is no longer the person designing the exam questions. This entails uncertainties on the part of both students and teachers.

Q2H2 formulates our expectation, as mentioned above, that there are no significant differences between teachers with and teachers without experience with statewide exams.

As to the individual self-efficacy (Bandura 1998) of teachers, we hypothesise that teachers who believe that they will be able to teach challenging students the subject matter covered in the exam and respond to students' needs individually will narrow teaching content less than teachers who do not hold this belief (Q2H3).

As the statewide exams in Hesse and Bremen are essentially a form of content standards, we address teacher cooperation with respect to the curriculum in our analysis as well. Cooperation among teachers is central to school development processes and to the successful implementation of policy reforms (e.g., Seashore

Louis, Marks & Kruse 1996). We expect that discussion about the curriculum facilitates its adaptation to instruction, but that this does not necessarily entail a narrowing of content. It may do so, however, if there is great insecurity among teachers. We therefore subject this variable to exploratory analysis without hypothesis.

At the school level, we consider the state as a confounding variable without formulating a hypothesis. Furthermore, we assume that students' abilities can induce teachers to widen or narrow the content. Teachers who expect their students to pass the statewide exam easily or with high marks may vary teaching content to a higher degree than teachers of students with heterogeneous achievement levels: The latter may, for example, focus on exam content to ensure that all of the students pass the exam. Our hypothesis is therefore that curricular content will be narrowed less at schools where students are high achievers or where they show homogeneous achievement levels than at schools where students do not perform as well or show heterogeneous achievement levels (Q2H4 and H5).

Other variables such as gender and teaching experience were tested prior to this analysis in exploratory studies, but these factors showed no influence.

### **Design and Sample**

In the DFG-funded study, questionnaires were distributed to students in two advanced and two basic-level courses in mathematics and English and to all teachers at 37 schools, both before and after the exit exams in 2007 and 2008. The students were also given a cognitive abilities test (CAT) (Heller & Perleth 2000).

In Hesse and Bremen, statewide exams were introduced by the government at different times and in different subjects; the schools themselves had no influence on

this process. In Hesse, statewide exams were introduced in 2007 and implemented simultaneously in all three written exam subjects. In Bremen, they were introduced in two stages: in 2007 in only the third exam subject (that is, in the basic-level course) and in 2008 in the advanced-level core subjects (German, mathematics, natural sciences, advanced foreign languages). In the other advanced-level courses, class-based testing continued after 2008. In Bremen, it is possible to compare teachers' responses to class-based exams with teachers' responses to statewide exams. Course-specific or subject-specific analyses are unavailable due to data protection.

As Bremen is a small state, our survey includes all schools (N=19). In Hesse, which is much bigger than Bremen, we used a representative sample of schools (N=18). Table 2 shows the survey response rates, which can be considered adequate (minimum: 47%, maximum: 72%).

**\*\*\*Table 2. Survey response rates in annual comparison\*\*\***

**\*\*\* about here \*\*\***

## **Instruments**

In the following, all of the variables—both dependent and independent—are explained. First, the dependent variable is described in detail; then, at least one item example is given for the remaining variables. For most of the variables, teachers were invited to rate the following statements from one to four (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree) with the exception of cooperation, which was coded from one to five (stated below). The items were then averaged into a scale. Variables coded differently are mentioned below.

To operationalise TTT, we measured the opposite of this phenomenon: content variance. This is defined as the dependent variable in both research questions. We asked teachers to rate their (dis-) agreement with the following five statements.

- “I took the desires and interests of my students into account in choosing the topics/content of the course.”
- “I took time to present topics of current importance in choosing the topics/content of the course.”
- “I took my own desires and interests into account in choosing the topics/content of the course.”
- “Although the statewide exams have pre-defined topics, I was able to cover my own topics as well.”
- “I also covered topics that were not part of the statewide exit exams.”

(Bremen: Cronbach’s  $\alpha_{2008 \text{ class-based}} = .84$ ,  $\alpha_{2008 \text{ statewide}} = .79$ ; Hesse: Cronbach’s  $\alpha_{2008 \text{ statewide}} = .77$ ). High scores on the content variance scale show that the teachers considered not only exam content but also, for example, students’ interests or topics of current relevance in designing their class curriculum. In Bremen, teachers of classes with class-based and statewide exams were invited to rate these statements. We therefore have two answers for teachers from Bremen concerning the variable “content variance.”

To measure teacher cooperation (Cronbach’s  $\alpha_{\text{Hesse}} = .89$ ,  $\alpha_{\text{Bremen}} = .90$ ), we modified a scale of Steinert, Gerecht, Klieme, and Döbrich (2003) and Halbheer, Kunz, and Maag Merki (2005), asking: “How often do you meet with other teachers from your school to discuss the following topics or problems?” One of the five topics, for example, was lesson preparation. Teachers were asked to state whether they met once a week, monthly, several times a year, once a year, or not at all.

The individual self-efficacy of teachers was measured with six items based on the scale by Schwarzer and Jerusalem (1999) (Cronbach's  $\alpha_{\text{Hesse}} = .73$ ,  $\alpha_{\text{Bremen}} = .72$ ), for example: "I know that I can manage to teach the material covered in the exam to even the most challenging students."

Uncertainty was measured with four items (Cronbach's  $\alpha_{\text{Hesse}} = .69$ ,  $\alpha_{\text{Bremen}} = .74$ ), which include: "I'm afraid that a topic will come up that the students aren't well prepared for."

Participation in the statewide exams is a dummy variable and was coded as 0 = no participation in statewide exams during the previous school year, 1 = participation in statewide exams during the previous school year.

Experience with statewide exams is also a dummy variable, and was coded as follows: If the teacher never taught a class with a statewide exam, he or she was assigned the value zero. If he or she taught a class with a statewide exam once or twice, he or she was assigned the value one or two.

The variable for the state is again a dummy variable (0 = Hesse, 1 = Bremen).

In the cognitive ability test (CAT), students could achieve a maximum score of 25 points, whose mean and standard deviations for each school were used in the analysis.

## **Analysis**

For the first research question, we evaluated the 2007 and 2008 data (collected prior to the exit exam) descriptively and carried out a comparison of teachers' responses from Bremen for statewide exams and class-based exams, respectively. We compared the mean scores using a t-test and a Cohen's d. Double responses (for statewide and class-based exams) were both considered in these analyses.

Moreover we conducted a multilevel analysis using the Hierarchical Linear and Nonlinear Modeling (HLM) software package by Raudenbush and Bryk et al. (2004). We analysed the 2008 data, attempting to control for years of experience with and current participation in statewide exams and for the state. We selected 2008 because this was the second year of statewide exams in both states, offering a situation in which teachers had already been given a chance to adapt to the new testing instrument. In this analysis, which is concerned with a few variables for which we have, of course, only one answer, it was not possible to consider double responses of teachers affected by both class-based and statewide exams in the variable “content variance.”

Therefore, preceding our analysis, we tested for differences in response behaviour between the two groups of teachers: the group of teachers who had experienced both types of exams during one school year and therefore responded twice to our variable content variance (Group A) and the group of teachers who had experienced just one of the two exam types during one school year (Group B). We found that teachers in Group A rated the variable content variance for the subjects with statewide exams at a similar level to the teachers in Group B. This was not the case when we compared the ratings for the subjects with class-based exams. Here, the results show that teachers in Group A had a significantly higher perception of having increased content variance than the teachers in Group B. To test our hypothesis in a conservative manner, we decided, with respect to the double responses of teachers in Bremen who rated the variable “content variance” for statewide as well as for class-based exams, to consider only the responses for statewide exams and not those given for the class-based exams ( $N = 1162$ ,  $M = 2.35$ ,  $SD = 0.60$ ). This ensures that we do not bias our results in the direction of our hypothesis.



At the individual level, we considered participation in the statewide exams in 2008 ( $N = 1162$ ,  $M = 0.92$ ,  $SD = 0.26$ ) and the years of experience with statewide exams ( $N = 1655$ ,  $M = 1.06$ ,  $SD = 0.80$ ). At the school level, we considered the state ( $N = 37$ ,  $M = 0.51$ ,  $SD = 0.51$ ). With the exception of the variable “experience with the statewide exams” (grand-mean centred), we did not centre the variables in our analysis. We analysed the following multilevel regression equation, in which interaction effects were not included:

$$\text{Content Variance} = \gamma_{00} + \gamma_{01} * \text{Federal State} + \gamma_{10} * \text{Experience} + \gamma_{20} * \text{Participation} + u_0 + r$$

To answer the second research question, we conducted a multilevel analysis of the 2008 data from Hesse and Bremen collected prior to the exit exam and only from teachers who were affected by the statewide exams in the current school year. We again used the software package HLM by Raudenbush and Bryk et al. (2004). The variables presented in Table 3 were included in the analysis:

**\*\*\*Table 3. Variables included in the multilevel analysis\*\*\***

**\*\*\* about here \*\*\***

With the exception of the variable “state,” each variable was grand-mean centred. In the following, we analyse the multilevel regression equation without interaction effects:

$$\begin{aligned} \text{Content Variance} = & \gamma_{00} + \gamma_{01} * \text{CAT median} + \gamma_{02} * \text{CAT standard deviation} \\ & + \gamma_{03} * \text{Federal state} + \gamma_{10} * \text{Cooperation} + \gamma_{20} * \text{Individual self - efficacy} \\ & + \gamma_{30} * \text{Uncertainty} + \gamma_{40} * \text{Experience} + u_0 + r \end{aligned}$$

## **Results**

Tables 4 and 5 show the results for the descriptive statistics. In Bremen, teachers in the subjects with class-based exams “agreed” in 2007 and 2008 with the statement that they, for instance, had covered topics of current relevance (see Table 4: 2007:  $M = 2.94$ ; 2008:  $M = 2.93$ ). However, the teachers of subjects with statewide testing responded on average that they “disagreed” with this statement (Table 4: 2007:  $M = 2.38$ ; 2008:  $M = 2.40$ ).

**\*\*\*Table 4. Bremen: Content variance in subjects with class-based and statewide testing (teacher statements)\*\*\***

**\*\*\* about here \*\*\***

A t-test comparing scores for teachers in the class-based exam system and in the statewide exam system showed highly significant differences that are, due to the effect size, relevant in both years ( $p < .01$ ; 2007: Cohen's  $d = -1.01$ ; 2008: Cohen's  $d = -0.94$ ). Cohen's  $d$  is defined as the difference between two means divided by their standard deviation ( $d < |.20|$  is defined as a small effect,  $d \sim |.50|$  as a medium effect and  $d > |.80|$  as a large effect). The results from Hesse, where there is only a statewide testing system, confirm those from Bremen (see Table 4 and 5): Teachers of subjects with statewide exams perceived a greater reduction of content (2007:  $M = 2.29$ ; 2008:  $M = 2.24$ ) than the Bremen teachers of subjects in which class-based exams were being administered. In Hesse, the absolute score for subjects with statewide exams was even below that in Bremen for subjects with statewide exams.

**\*\*\*Table 5. Hesse: Content variance in subjects with statewide testing (teacher statements).\*\*\***

**\*\*\* about here \*\*\***

The results of the multilevel analysis (see Table 6), in which we controlled for participation in and experience with statewide exams and for the state show that the school level has little impact on the degree to which teachers reduce curricular content (5% variance can be explained (result of the intra-class correlation (0.5), which is the proportion of variance attributed to schools)). Experience with statewide exams had no influence on the narrowing of curricular content by teachers. It did make a difference, however, whether a teacher was currently participating in statewide exams. That is, teachers who were currently participating in statewide exams limited content variance more than teachers who were not. Moreover, teachers from Bremen narrowed the content to a lower degree than teachers from Hesse.

**\*\*\*Table 6. Statewide and class-based exam data from two German states**

**(2008): Content variance.\*\*\***

**\*\*\* about here \*\*\***

To answer the second research question (What factors are responsible for teaching to the test?), we again conducted a multilevel analysis, but this time focused exclusively on teachers of classes with statewide exams. Table 7 presents the results. Model 1 is the intercept-only model, Model 2 includes the individual-level variables, and Model 3 shows the entire model (individual- and school-level variables).

Just three percent of the variance can be explained at the school level (result of the intra-class correlation (0.3)). The results in Models 2 and 3 show that high individual self-efficacy can prevent teachers from reducing content variance

(standardised coefficient [stand. coeff.] = 0.17). However, teachers' uncertainty with regard to the requirements of statewide exams (stand. coeff. = -0.16) is correlated with their focus on adapting the curriculum to exam content and therefore leaving topics of current relevance aside. Prior experience with statewide exams and cooperation among teachers do not show significant effects.

At the school level, neither the average scores on the cognitive abilities tests nor the standard deviations were able to explain the variance (see M3). However, there is a correlation between the state and the variable of teachers reducing content variance (stand. coeff. = 0.14). The TTT effect in Bremen is smaller than in Hesse (cf. M3).

**\*\*\*Table 7. Statewide exam data from two German states (2008): Content variance\*\*\***

**\*\*\* about here \*\*\***

## **Discussion**

*Do the statewide low-stakes exit exams produce the effect of teaching to the test from the teachers' point of view?*

The first research question must be answered in the affirmative (Q1H1). That is, even statewide exams that make no use of external examiners and do not bear potentially severe consequences for either schools or teachers still involve teaching methods that ignore students' interests and relevant current issues. We assumed that the introduction of standards that can be described essentially as content standards might contribute to this effect. Moreover, the fact that the stakes introduced with statewide exams are low—but still higher overall than they were before the implementation of the statewide exams—could be a reason for this result.

Moreover, our analysis shows a difference between current participation in statewide exams and experience with statewide exams. There is evidence for both

hypotheses (Q1H2 and Q1H3): Teachers who are participating in statewide exams in the current school year narrow the curriculum to a higher degree than teachers who are not. Teachers with two years of experience with statewide exams reduce content variance to the same degree as teachers with one year of experience. This shows that teachers are unable to profit from the learning effect of prior experience with statewide exams because new topics are introduced every year.

However, a few limitations of our study have to be considered. Due to data protection, we were unable to collect and analyse subject- or class-specific information. Moreover our results are limited because causal relationships could not be identified. In future studies, it would be interesting to analyse student data as well to study the issue of content variance from this additional perspective.

*What factors, aside from the use of statewide exit exams, are responsible for teaching to the test?*

With regard to our second research question, our findings confirm the first hypothesis (Q2) that high uncertainty about the requirements of the statewide exams leads to a greater reduction of content variance than low uncertainty. Here, further studies are needed to investigate how teachers' uncertainty changes over time.

We also find evidence confirming our second hypothesis (Q2) that due to particular features of statewide exams, which create information asymmetries between teachers and the institutions designing the exam questions, teacher experience with statewide exams is not able to counteract the narrowing of content.

Moreover, the multilevel analysis shows that teachers' perception of individual self-efficacy goes hand in hand with a tendency to cover current topics and take student needs into account (Q2H3).

In addition, our analysis shows that cooperative behaviour has no impact on the narrowing or widening of content variance. This is partly in line with our hypothesis that cooperation should only have an effect if it coincides with additional conditions. These conditions, which include uncertainty in the teaching staff (as a whole), are not considered in our analysis.

Although very little of the variance can be explained at the school level, it makes a difference whether a school is located in Hesse or Bremen. The origins of this effect (social, cultural, details of test design, etc.) cannot be explained in the context of this study. Teachers in Bremen report lower perceptions of content narrowing than teachers in Hesse, however.

Hypotheses 4 and 5 (Q2) proposed with respect to a possible effect of teachers' perceptions of students' cognitive abilities in the schools (CAT test: mean and standard deviation) were not confirmed by the results. This may be due to some critical aspects of the variable: the cognitive abilities test does not measure curricular requirements, nor does it take comparisons to social or individual benchmarks into account. Here, studies are needed in which teachers are asked to assess the heterogeneity or homogeneity of the student population directly. Moreover, the variable is analysed at the school level, although a class-based indicator would be more appropriate (this was not possible due to data protection).

Since the TTT effect was defined at the outset of this study as a narrowing of content, a tendency to leave aside topics relevant to everyday life, and a failure to consider student interests, we can confirm that the statewide exams implemented in Hesse and Bremen in 2008 correlate—in the perceptions of teachers—with these unintended effects. Our study is not equipped to analyse whether this reduces student motivation and leads in turn to lower performance. Future studies will have to

compare findings on the TTT effect to results on learning strategies, subject interest, and exit exam or CAT test results. In addition, longitudinal analyses should be carried out to check for effect consistency. Moreover, our definition of the TTT effect does not address the questions of whether classroom assessments mirror the format, item type, or cognitive demands of the statewide exams as described by Abrams (2007). Nor does it include the reallocation of instructional time across grades or subjects as pointed out by Hamilton & Stecher (2002).

## **Conclusion**

In summary, we conclude that a TTT effect does appear, even under conditions that allow teachers professional discretion in curricular design. In theory, such discretion might weaken the “crowding-out” effect. That is, less external intervention prevents intrinsic motivation from being undermined and thus affects the agent’s marginal benefit from performing (Frey & Jegen: 2000). The fact that we still observe a narrowing of curricular content might be caused by the introduction of content standards and the fact that the specific type of exams introduced in Hesse and Bremen is exit exams. It seems that teachers use their students’ exam results as a form of personal feedback. Teachers want to prepare their students as best as they can; we assume that their strategy for improving their students’ results in the statewide tests is to decrease content variance. But unfortunately, they tend to narrow curricular content to an extent that might create problems for their students and for their own work as well. Based on our results so far, it can be assumed that teachers’ additional experience with statewide exams will not counteract this tendency. The curriculum narrowing identified here may not occur *despite* the degree of freedom granted to teachers in curricular design, but indeed *because* teachers are using their professional discretion. This may be seen as an indication that teachers feel accountable for their

students' performance and for reaching achievement goals and curriculum standards without external incentives. An alternative to statewide exit exams may therefore be to implement "opportunity-to-learn standards" throughout schooling instead of content standards at the end of school. If provided in connection with other forms of support equipping teachers to better interpret student learning outcomes and to find appropriate solutions for "failing" situations, opportunity-to-learn standards could act as a professional feedback system that could improve teaching and learning and avoid TTT effects. Further studies are needed on this subject.

## References

- Abrams, L. & Madaus, G. (2003). The lessons of high-stakes testing. *Education. Leadership*, 61(3), 31-35.
- Abrams, L. M. (2007). Implications of high-stakes testing for the use of formative classroom assessment. In McMillan, J. H. (Ed.) *Formative classroom Assessment*. Danvers, USA: Teacher College, Columbia.
- Alderson, J. Charles, & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115–129.
- American Evaluation Association (Ed.) (2009). Position statement on high stakes testing in PreK-12 Education. Retrieved June 3, 2009 from <http://www.eval.org/hst3.htm>, last updated on June 3, 2009.
- Andrews, S. (2004). Washback and curriculum innovation. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: Research contexts and methods* (pp. 37–50). Mahwah, N.J., London: Lawrence Erlbaum.
- Ardalan, A., Quarstein, V. A., & McAfee, R. Bruce (1994). Enhancing performance through employee discretion and feedback. *Industrial Management & Data Systems*, 94(10), 15–19.
- Au, W. (2007). High stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36, 258.
- Bandura, A. (1998). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bishop, J. H. (1995). The impact of curriculum-based external examinations on school priorities and student learning. *International Journal of Educational Research*, 23(8), 653–752.
- Brozo, W. G. & Hargis, C. (2003). Using low-stakes reading assessment. *Educational Leadership*, 61(3), 60–64.
- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng & Y. Watanabe (Eds.), *Washback in*



*language testing: Research contexts and methods* (pp. 3–19). Mahwah, N.J., London: Lawrence Erlbaum

Cheng, L., & Watanabe, Y. (Eds.) (2004). *Washback in language testing: Research contexts and methods*. Mahwah, N.J., London: Lawrence Erlbaum.

Clarke, M., Haney, W., & Madaus, G. (2000). High stakes testing and high school completion. NBETPP statements, The National Board on Educational Testing and Public Policy, 1(3), 1–11.

Corbett, H. D. & Wilson, B. L. (1991). Testing, reform, and rebellion. Norwood, NJ: Ablex (Interpretive perspectives on education and policy).

Deci, E. Lewis, Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(3), 627–668.

EU Commission (2009). The Lifelong Learning Programme: education and training opportunities for all. Internet source: [http://www.lebenslanges-lernen.eu/leonardo\\_da\\_vinci\\_4.html](http://www.lebenslanges-lernen.eu/leonardo_da_vinci_4.html) (retrieved on November 30, 2009)

Frey, B. S., & Jegen, R. (2000). *Motivation crowding theory: A survey of empirical evidence* (Revised version, Vol. 49). Working paper / Institute for Empirical Research in Economics, 49. Zurich: University of Zurich, Institute for Empirical Research in Economics.

Glaser, J., Herbig, B., & Gunkel, J. (2008). *Kreativität und Gesundheit im Arbeitsprozess: Bedingungen für eine kreativitätsförderliche Arbeitsgestaltung im Wirtschaftsleben*. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.

Hacker, W. (2005). *Allgemeine Arbeitspsychologie: Psychische Regulation von Wissens-, Denk- und körperlicher Arbeit* (2., vollst. überarb. und erg. Aufl., 58/2). Schriften zur Arbeitspsychologie, 58/2. Bern: Huber.

Halbheer, U., Kunz, A., & Maag Merki, K. (2005). *Pädagogische Entwicklungsbilanzen an Zürcher Mittelschulen. Indikatoren zu Kontextmerkmalen gymnasialer Bildung. Perspektive der Lehrpersonen: Schul- und Unterrichtserfahrungen. Skalen- und Itemdokumentation*. Zürich: Forschungsbereich Schulqualität & Schulentwicklung, Pädagogisches Institut, Universität Zürich.

Hamilton, L. S. & Stecher, B. M. (2002). Improving test-based accountability. In Hamilton, L. S., Stecher, B. M., & Klein, S. P. (Eds.): Making sense of test-based accountability in education (pp. 121–144). Santa Monica, CA: RAND Education.

Hamilton, L. S., Stecher, B. M., Russell, J. L., Marsh, J. A., & Miles, J. (2008). Accountability and teaching practices: School-level actions and teacher responses. In Fuller, B., Henne, M. K., & Hannum, E. (Eds.): Strong states, weak schools: The benefits and dilemmas of centralized accountability (pp. 31–66). Research in Sociology of Education, 16.

Heller, K. A., & Perleth, C. (2000). Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4-12+ R). Göttingen: Beltz Test.

Hoppe, C. (2004). Die Fortbildungsoffensive. In Projektleitung “Selbständige Schule” (Ed.) Verantwortung für Qualität – Grundlagen des Projektes, 1, (pp. 75-83). Troisdorf: Bildungsverlag EINS.

- Jordan, P. C. (1986). Effects of an extrinsic reward on intrinsic motivation: a field experiment. *Academy of Management Journal*, 29(2), 405–412.
- Klein, E. D., Kühn, S. M., van Ackeren, I., & Block, R. (2009). Wie zentral sind zentrale Prüfungen? Abschlussprüfungen am Ende der Sekundarstufe II im nationalen und internationalen Vergleich. *Zeitschrift für Pädagogik*, 55(4), 596–621.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., & Prenzel, M. et al (2003). Zur Entwicklung nationaler Bildungsstandards: Eine Expertise (3., unveränderte Aufl., Band 1). Bildungsreform, Band 1. Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- KMK - Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005). Bildungsstandards der Kultusministerkonferenz: Erläuterungen zur Konzeption und Entwicklung. München: Wolters Kluwer Deutschland.
- Liermann, C. (2009). Individualkonzepte unter dem Einfluss von Standardisierungsprozessen in der Schule. In J. Bilstein & J. Ecarius (Eds.), *Standardisierung - Kanonisierung: Erziehungswissenschaftliche Reflexionen* (1st ed., pp. 235–251). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Madaus, G. F. (1988). The influence of testing on the curriculum. In L. N. Tanner (Ed.), *The yearbook of the National Society for the Study of Education: . Critical issues in curriculum* (pp. 83–121). Chicago, Ill.: Univ. of Chicago Pr.
- Niskanen, W. A. (1971). *Bureaucracy and representative government* (1. publ.). Chicago: Aldine Atherton.
- OECD. (2004). Bildung auf einen Blick – OECD-Indikatoren 2004. Deutschland 2004. Bundesministerium für Bildung und Forschung.
- Oelkers, J. (2007). Bildungsstandards und Individualförderung: Ein Widerspruch? Lecture to the Verband der Schulpflege-Präsidentinnen und -Präsidenten Luzern (VSPL) on October 22, 2007, in Emmenbrücke. Retrieved on September 23, 2009, from [http://paed-services.uzh.ch/user\\_downloads/601/291\\_Emmen bruecke.pdf](http://paed-services.uzh.ch/user_downloads/601/291_Emmen_bruecke.pdf).
- Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Raudenbush, S.W., Bryk, A.S., & Congdon, R. (2004). HLM 6 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.
- Rhoades, K. & Madaus, G. (2003). Errors in standardized tests: a systematic problem: Boston, MA: Boston College.
- Richter, I. (2004). Chancen und Risiken selbständiger Schulen in Nordrhein-Westfalen. In Projektleitung "Selbständige Schule" (Ed.): Verantwortung für Qualität – Grundlagen des Projektes, 1, (pp. 1-29), Troisdorf: Bildungsverlag Elns.
- Schwarzer, R. & Jerusalem, M. (Hrsg.). (1999). *Skalen zur Erfassung von Lehrer und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin und Humboldt-Universität zu Berlin.

Seashore Louis, K., Marks, H., & Kruse, S. (1996). Teachers' Professional Community in Restructuring Schools. *American Educational Research Journal*, 33(4), 757–798.

Steinert, B., Gerecht, M., Klieme, E. & Döbrich, P. (2003). *Skalen zur Schulqualität: Dokumentation der Erhebungsinstrumente. Arbeitsplatzuntersuchung (APU), Pädagogische Entwicklungsbilanzen (PEB)*. Frankfurt a.M.: Dipf, Deutsches Institut für Internationale Pädagogische Forschung.

Ulich, E. (1994). *Arbeitspsychologie* (3. Aufl.). Zürich: vdf.

Valli, L., Croninger, R. G., Chambliss, M. J., Graeber, A. O., & Buese, D. (2008). *Test driven. High-stakes accountability in elementary schools*. New York: Teachers College Press.

Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: Research contexts and methods* (pp. 19–36). Mahwah, N.J., London: Lawrence Erlbaum.

## Tables

**Table 1: Instruments that may affect the TTT effect**

Five factors mediating the effect of washback (Watanabe 2004: 22)	Eleven specified aspects (cf. literature)
Macro context factors	Content standards (Klieme et al. 2003) Performance standards (Klieme et al. 2003)
Micro context factors	External examiner (Klein et. al. 2009) Anonymous grading (Klein et. al. 2009)
Personal factors	Personal factors (Abrams et al. 2003)
Prestige factors	Consequences for the school district (Pedulla, Abrams, Madaus, Russell, Ramos & Miao 2003) Consequences for the school (Pedulla et al. 2003) Consequences for the teacher (Pedulla et al. 2003) Consequences for the students (Pedulla et al. 2003) / Weight of the exam in the final grade at least 50% (Klein et al. 2009)
Test factors	Test design (Rhoades & Madaus 2003, Oelkers 2007) Number of exam subject (Klein et.al. 2009)

**Table 2. Survey response rates in annual comparison**

Year	Hesse		Bremen	
	Teacher	Students	Teacher	Students
2007	1092 (72.3%)	973 (67.5%)	641 (68.3%)	751 (46.6%)
2008	949 (59.1%)	975 (71.3%)	581 (64.2%)	977 (65.2%)

**Table 3. Variables included in the multilevel analysis. Data from 2008 with regard to teachers who were affected by statewide exams**

Model	Variable	N	M	SD	Min.	Max.
Dependent variable	Content variance	1058	2.30	0.60	1	4
Independent variables						
Individual level	Uncertainty*	1046	2.11	0.57	1	4
	Experience*	1629	1.06	0.80	0	2
	Individual self-efficacy	1439	2.84	0.44	1	4
	Cooperation curriculum	1457	2.76	0.90	1	5
School level	State	36	0.50	0.51	0	1
	CAT mean	36	16.25	1.43	13.17	18.72
	CAT standard deviation	36	4.18	0.78	2.70	7.36

Note: N = number of persons; M = mean; SD = standard deviation; \* with statewide exams

**Table 4. Bremen: Content variance in subjects with class-based and statewide testing (teacher statements)**

Scale		2007			2008		
		N	M	SD	N	M	SD
Content variance	Class-based	309	2.94	0.48	183	2.93	0.45
	Statewide	395	2.38	0.60	394	2.40	0.61

Note: Answer options: 1 = strongly disagree; 2 = disagree; 3 = agree; 4 = strongly agree; N = number of persons; M = mean; SD = standard deviation; p = significance level, n.s. = not significant

**Table 5. Hesse: Content variance in subjects with statewide testing (teacher statements)**

Scale	2007			2008		
	N	M	SD	N	M	SD
Content variance	766	2.29	0.57	681	2.24	0.58

Note: Answer options: 1 = strongly disagree; 2 = disagree; 3 = agree; 4 = strongly agree; N = number of persons; M = mean; SD = standard deviation; p = significance level

**Table 6. Statewide and class-based exam data from two German states (2008): Content variance**

Fixed part predictor	M1: Intercept-only		M2: M1 + Individual + school level variables	
	coeff. (SE)	p	coeff. (SE)	p
Intercept	2.37 (0.03)	< .001	2.79 (0.05)	< .001
Experience with statewide exams			0.01 (0.03)	n.s.
Participation in statewide exams			-0.54 (0.04)	< .001
State			0.16 (0.04)	< .01
Random part				
u <sub>00</sub>	0.02	< .001	0.00	< .05
r <sub>0</sub>	0.35		0.33	
Statistics for current covariance components model				
Deviance	2103		2031	

Note: M = model; coeff = coefficient; SE = standard error; p = significance level; n.s. = not significant (p > .05); u<sub>00</sub> is the unique increment to the intercept associated with the schools after controlling for school level variables, r<sub>0</sub> represents the residual variance at level-1 after controlling for the individual level variables (see Raudenbush & Bryk et al. 2004)

**Table 7. Statewide exam data from two German states (2008): Content variance**

Fixed part predictor	M1: Intercept-only		M2: M1 + individual level variables		M3: M2 + school level variables	
	coeff. (SE)	p	coeff. (SE)	p	coeff. (SE)	p
Intercept	2.32 (0.02)	< .001	2.32 (0.02)	< .001	2.24 (0.03)	< .001
Uncertainty*			-0.17 (0.04)	< .001	-0.17 (0.04)	< .001
Experience*			-0.00 (0.03)	n.s.	0.01 (0.03)	n.s.
Individual self-efficacy			0.22 (0.04)	< .001	0.24 (0.04)	< .001
Cooperation curriculum			0.02 (0.02)	n.s.	0.03 (0.02)	n.s.
State					0.17 (0.04)	< .001
CAT mean					-0.00 (0.02)	n.s.
CAT standard deviation					-0.01 (0.03)	n.s.
Random part						
u <sub>00</sub>	0.01	< .01	0.01	< .01	0.00	n.s.
r <sub>0</sub>	0.34		0.32		0.32	
Statistics for current covariance components model						
Deviance	1899		1723		1727	

Note: M = model; coeff = coefficient; p = significance level; n.s. = not significant (p > .05); u<sub>00</sub> is the unique increment to the intercept associated with the schools after controlling for school level variables, r<sub>0</sub> represents the residual variance at level-1 after controlling for the individual level variables (see Raudenbush & Bryk et al. 2004); \*with statewide exam

## Acknowledgements

---

<sup>1</sup> The OECD countries are Denmark, Finland, France, Ireland, Italy, Luxemburg, New Zealand, the Netherlands, Norway, Poland, Portugal, Scotland, Slovakia, Hungary, and the UK.

<sup>2</sup> See

[http://www.kultusministerium.hessen.de/irj/zentral\\_Internet?rid=zentral\\_15/zentral\\_Internet/nav/bd7/bd75058e-6897-0701-33e2-dc44e9169fcc,f9b20903-2748-7221-b30b-cd44e9169fcc,22222222-2222-2222-2222-222222222222,22222222-2222-2222-2222-222222222222,11111111-2222-3333-4444-100000005004.htm](http://www.kultusministerium.hessen.de/irj/zentral_Internet?rid=zentral_15/zentral_Internet/nav/bd7/bd75058e-6897-0701-33e2-dc44e9169fcc,f9b20903-2748-7221-b30b-cd44e9169fcc,22222222-2222-2222-2222-222222222222,22222222-2222-2222-2222-222222222222,11111111-2222-3333-4444-100000005004.htm) (retrieved on September 8, 2009).

<sup>3</sup> See [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2008/2008\\_10\\_24-VB-EPA.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_10_24-VB-EPA.pdf) (retrieved on January 21, 2011).

Daniela J. Jäger, 2012: Herausforderung Zentralabitur: Unterrichtsinhalte variieren und an Prüfungsthemen anpassen. In: Maag Merki, Katharina (Hrsg.): *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland*. Wiesbaden: Springer VS, S. 179-205.

Hinweis

The final publication is available at  
[http://link.springer.com/chapter/10.1007/978-3-531-94023-6\\_8](http://link.springer.com/chapter/10.1007/978-3-531-94023-6_8).

## **8. Herausforderung Zentralabitur: Unterrichtsinhalte variieren und an Prüfungsthemen anpassen**

*Daniela J. Jäger*

### **1 Ausgangslage**

Das Zentralabitur verlangt von den Lehrpersonen, den Schüler/-innen den Stoff zu vermitteln, der in den Schwerpunktthemen für das Abitur seitens des Ministeriums formuliert wurde und gegebenenfalls im Abitur abgefragt wird. Dies klingt banal, verändert aber die Rolle der Lehrenden grundlegend. Denn die Einheit von Unterrichtenden und Prüfenden wird zugunsten der Vergleichbarkeit der Schüler/-innenleistungen aufgebrochen. Eine vermeintlich objektive, dritte Instanz soll die Leistungen der Schüler/-innen und – indirekt auch – die der Lehrenden messen. Diese erhöhte Kontrolle kann die Selbstbestimmung der Lehrenden soweit einschränken, dass es zu unerwünschten Effekten, wie dem so genannten Teaching-to-the-Test(ttt)-Effekt, kommt. Das bedeutet, dass in der negativen Ausprägung der Unterricht u.U. mehr als erforderlich auf die inhaltlichen Vorgaben der zentralen Prüfungen beschränkt wird (vgl. z.B. Stecher, 2002).

In diesem Artikel wird untersucht, inwiefern die Lehrpersonen in Hessen und Bremen die Herausforderung, die Interessen der Schüler/-innen nicht zu vernachlässigen und außerdem Kongruenz zwischen Unterricht und Prüfung herzustellen, aus eigener Sicht in den ersten Jahren nach der Implementation des Zentralabiturs meistern. Zudem wird analysiert, welche Faktoren dazu beitragen, einem negativen ttt-Effekt vorzubeugen bzw. Kongruenz zwischen Unterricht und Abiturprüfung herzustellen.

### **2 Theoretischer Hintergrund**

#### *2.1 Herausforderung Zentralabitur?*

Nach dem TIMSS- und PISA-Schock wurde seitens der Regierungen u.a. ein Steuerungsdefizit im Bildungssektor diagnostiziert (vgl. z.B. Steffens, 2007). Mittels Instrumenten der New Public Management(NPM)-Theorie sollten die Bildungserzeugenden stärker kontrolliert werden. Das bedeutet, die Kontrolle sollte erhöht werden, um sicherzustellen, dass die Lehrenden den Vorgaben z.B. zum Abitur folgen, dass sie kongruent zu den Vorgaben handeln. Um dies zu er-



reichen, wurde die Rolle der Lehrenden in Lehrende und Prüfende aufgebrochen. Gleichzeitig sollten bei den Reformen die Bildungsangebote besser an die Bedürfnisse der „Konsument/-innen“ angepasst werden.

Die NPM-Theorie beruht auf Annahmen aus der Wirtschaft. Hier sollen Aufgaben in Firmen durch „Outsourcing“ bzw. „Contracting Out“ effizienter organisiert werden (vgl. u.a. Bogumil & Jann, 2009; Schedler, 1995). Zudem wird dem daraus folgenden Problem der Informationsasymmetrie zwischen Prinzipal (Auftraggeber) und Agent (Ausführendem) durch die Einführung von Anreizsystemen entgegengewirkt („principal-agent-theory“, Niskanen, 1971). Trotz der Schwierigkeiten (vgl. z.B. Weiß, 2003), diese Theorie von Firmen aus dem privaten Sektor auf Institutionen, die öffentliche, wenn nicht gar meritokratische Güter bereitstellen, zu übertragen, wurde die Idee des NPMs auch im Bildungswesen weitgehend übernommen (OECD, 2004).

Die Idee beinhaltet, den Erfolg von Institutionen und Unternehmen anhand von Outputfaktoren zu messen. Voraussetzung zur Anwendung des NPMs sind also vergleichbare Indikatoren. Diese werden zwischen dem Prinzipal und dem Agenten vertraglich festgelegt (Kontraktmanagement). Die Einhaltung des Vertrages, die erfolgreiche oder missglückte Umsetzung der festgelegten Indikatoren, wird vom Prinzipal sanktioniert bzw. honoriert. Gleichzeitig soll die Umsetzung der Outputfaktoren veröffentlicht werden, um so Wettbewerb zwischen den Anbietern (Agenten) zu erzeugen („Management by Competition“, Bogumil & Jann, 2009, S. 242). Dies soll einerseits dazu führen, dass der Prinzipal nur Verträge mit den effizientesten Anbietern abschließt, andererseits soll so auch die Marktmacht der Konsument/-innen gestärkt werden. Die Nachfrage – so die Theorie – soll das Angebot regeln. Im Zeitverlauf soll die Leistung der Anbietenden gesteigert werden.

Analog sind im Bildungssystem die Indikatoren für das Outputmanagement beispielsweise die Ergebnisse der zentral vorgegebenen Abschlussprüfungen. Auf ein „Management by Competition“ wird in Deutschland – im Gegensatz beispielsweise zu den USA – verzichtet. In Deutschland wird daher auch von einem Low-stakes- und in z.B. den USA von einem High-stakes-System gesprochen. So werden die Ergebnisse der Leistungen der Schüler/-innen einer Schule nicht der Öffentlichkeit mitgeteilt und sie sind auch nicht an Honorare bzw. Sanktionen für Lehrpersonal und/oder Schulen gebunden. Allerdings können in einigen Bundesländern die Eltern und Schüler/-innen eine Schule auswählen. Auf diese Weise wird bereits Wettbewerb unter den Schulen erzeugt. Zudem werden die Ergebnisse der Abiturprüfungen den Lehrkräften einer Schule mitgeteilt. So wird indirekt ein privater „feedback loop“ (Hamilton, Stecher, Russell, Marsh & Miles, 2008) installiert. Dieser soll die Lehrpersonen dazu anregen, den Unterricht

besser auf die im Abitur verlangten Inhalte abzustimmen. Die Kongruenz zwischen Unterricht und zentral gestellter Prüfung soll erhöht werden.

Im Sinne des NPMs sollte sich über die Jahre das Angebot (Unterricht) an die Nachfrage (Inhalte der Abiturprüfung) anpassen. Da jedoch das Bildungssystem nicht mit Firmen identisch ist, hinkt u.a. hier die Übertragung der Theorie: Im Bildungssystem werden jedes Jahr neue, andere Abituraufgaben gestellt. Zudem ist zwar durch die Festlegung der Schwerpunktthemen „vertraglich“ geregelt, welche Inhalte den Schüler/-innen vermittelt werden sollen, die Angaben beinhalten aber keine Details zu der zu erbringenden – da als Output gemessenen – Leistung. Das heißt, die Lehrpersonen können nicht aufgrund des „feedback loop“ lernen, sondern sind jedes Jahr wieder der Unsicherheit ausgesetzt, ob der Unterricht den Erfordernissen der Aufgabenkommission gerecht werden wird.

Noch eine weitere Schwierigkeit ist mit der Einführung des Zentralabiturs als Steuerungsinstrument verbunden: Die Gefahr der „Übergeneralisierung des pay-for-performance-Prinzips“ (Frey & Jegen, 2002, S. 30). Mit Übergeneralisierung ist gemeint, dass der Agent nur Aufgaben erfüllt, die vertraglich zwischen ihm und dem Auftraggeber festgelegt wurden. Schwer messbare und damit nicht direkt bezahlte Aufgabengebiete werden dagegen vom Agenten vernachlässigt (ebd.). Diese Übergeneralisierung kann eine Reaktion auf den externen Eingriff durch monetäre Anreize oder Befehle sein: „Unter gewissen Bedingungen wird beim Versuch, intrinsisch motivierte Mitarbeiter/-innen oder Helfer/-innen mit zusätzlichen Geldleistungen weiter zu motivieren, der wertvolle Eigenantrieb reduziert oder sogar ganz zerstört“ (ebd., S. 31). Dieser Effekt, dass extrinsische Motivation die intrinsische verdrängt, wird in der „Motivation Crowding Theory“ als Verdrängungseffekt bezeichnet (ebd., S. 31). Bedingung für sein Auftreten ist erstens, dass intrinsische Motivation zuvor bei den Mitarbeitenden vorhanden ist, wovon bei den allermeisten Arbeiten ausgegangen wird, und zweitens, dass die intrinsische Motivation nur dann durch externe Interventionen unterwandert wird, wenn dadurch die Selbstbestimmung und Selbstachtung der Mitarbeitenden eingeschränkt werden (ebd., S. 32). Dies bedeutet für Unternehmen und Institutionen, dass externe Eingriffe die Selbstbestimmung der Mitarbeitenden möglichst wenig einschränken und zudem unterstützend statt kontrollierend eingesetzt werden sollten (ebd., S. 32). Frey und Jegen (2002, S. 34f) weisen zudem daraufhin, dass der Verdrängungseffekt besonders in Organisationen auftritt, die Gemeingüter bereitstellen. Denn hier können die Verträge mit den Beschäftigten aufgrund der Komplexität des Angebots nie alle relevanten Aspekte erfassen (ebd., S. 35). „Aus diesem Grund führen extensive Kontrollen oder Anreizkontrakte häufig zu unerwünschtem und für die Firma schädlichem Verhalten. Die Mitarbeitenden konzentrieren sich ausschließlich auf die monetär

belohnten Tätigkeiten, vernachlässigen hingegen diejenigen, für die sie kein monetäres Entgelt erhalten“ (ebd., S. 35).

Frey und Jegen (2002, S. 37) gehen weiter davon aus, dass die intrinsische Motivation auch indirekt verdrängt werden kann, und zwar durch Anreizsysteme in ähnlichen Bereichen, deren Bedeutung übertragen wird. Sie nennen diesen Effekt den Motivation-Transfer-Effekt. Als Beispiel führen sie ein Kind an, welches spielerisch den Rasen im elterlichen Garten mäht, dann aber dafür bezahlt wird. Tritt der Verdrängungseffekt auf, dann mäht das Kind nun nicht mehr ohne Lohn den Rasen. Zeigt sich der Motivation-Transfer-Effekt, dann fordert das Kind nun auch für andere Haushaltstätigkeiten Geld.

Übertragen auf das Bildungssystem bedeutet dies, dass durch den kontrollierenden Eingriff der Regierenden in die Rolle der Lehrenden als Unterrichtende und Prüfende die Selbstbestimmung in einer Weise tangiert wird, die einen Verdrängungseffekt auslösen kann. Die Themenvarianz im Unterricht wird damit von den Lehrenden auf die Themen im Abitur eingegrenzt, andere Arbeitsbereiche, die ebenfalls u.a. im Curriculum festgelegt sind, werden vernachlässigt.

## 2.2 *Forschungsstand*

Inwiefern es unter der Bedingung zentraler Prüfungen Lehrpersonen möglich ist, Kongruenz zwischen Unterricht und zentralen Prüfungen herzustellen, wird in der Literatur zumeist unter dem Aspekt „teaching-to-the-test“ diskutiert. Dabei wird der Begriff neutral verwendet. Die Einen betonen den positiven Aspekt, dass die Angaben des Ministeriums nun eher im Sinne des Ministeriums (Prinzipal) erfüllt würden, d.h. dass Kongruenz zwischen Unterricht und Prüfung erfolgreich hergestellt würde (Bishop, 1995; Hoppe, 2004; Richter, 2004). Die Anderen kritisieren die Übergeneralisierung der zuvor für das Abitur vorgegebenen Themen und beschreiben das ttt als reine Kopie der zentralen Prüfungen hinsichtlich der Unterrichtsinhalte, der Form des Unterrichtsgesprächs sowie den Anforderungen (vgl. z.B. Abrams, 2007; Au, 2007). Interessen der Schüler/-innen, Alltagsbezüge im Unterricht und Kompetenzen der Lehrpersonen würden vernachlässigt. Auch weitere Strategien werden in der Forschung zum ttt beobachtet, wie eine Verlagerung der Aufmerksamkeit auf die zentral geprüften Schulfächer oder aber auch auf die Schulstufen, in denen getestet wird (Hamilton & Stecher, 2002). Diese Strategien könnten bereits als Motivation-Transfer-Effekt interpretiert werden.

Die oben genannten Strategien sollen dazu dienen, die Kongruenz zwischen Unterricht und der zentral gestellten Abiturprüfung zu erhöhen. Inwiefern die Lehrpersonen mit diesen Strategien jedoch erfolgreich sind, ist fraglich. Den Unterricht ausschließlich auf die vermeintlichen Themen des Zentralabiturs zu kon-

zentrieren, kann auch bedeuten, den Unterricht auf Themen zu fokussieren, die letztlich nicht abgefragt werden. Im Unterricht vorgegebene Operatorenlisten zu trainieren, kann auch dazu führen, dass Schüler/-innen an Fragen die z.B. eigenes Reflektieren erfordern, scheitern. Unter der Bedingung zentraler Prüfungen sind Unterricht und Prüfung voneinander entkoppelt und damit bleibt auch die Unsicherheit darüber bestehen, ob eine Fokussierung auf die Vorgaben oder aber eine breite Interpretation der Vorgaben für die Bewältigung der Abituraufgaben hilfreich ist.

In der Literatur werden als Ursachen für das ttt die Unsicherheit gegenüber den zentralen Prüfungen (Liermann, 2009) sowie die Kategorisierung eines Schulsystems als „high stakes“ genannt (Abrams & Madaus, 2003; Brozo & Hargis, 2003). Unsicherheit entsteht durch die Unwägbarkeit der Fragen im Abitur. Dies kann, wie oben erwähnt, zu einer oberflächlichen Bearbeitung eines sehr ausgeweiteten Themengebiets oder aber zu einer starken Einschränkung des Unterrichtsstoffes führen (Liermann, 2009, S. 246f). *Beide* Verhaltensweisen können die Kongruenz zu den Abiturprüfungen abschwächen.

Low-stakes-Tests könnten einem ttt vorbeugen, so Abrams und Madaus (2003) sowie Brozo und Hargis (2003). Denn durch den Wegfall von Sanktionen soll der Druck auf die Lehrpersonen gesenkt werden, der für das ttt verantwortlich gemacht wird. In diesem Beitrag wird allerdings die These aufgestellt (siehe Kapitel 3 in diesem Beitrag), dass auch in zentralen Low-stakes-Prüfungssystemen ein ttt-Effekt auftreten kann. Die Argumentation ist, dass nicht nur monetäre Anreize oder Konkurrenzsituationen zu Übergeneralisierungen führen, sondern eben auch jeder externe Eingriff in Form eines Befehls (Frey & Jegen, 2002). Dies besonders in Einrichtungen, die gemeinnützige Güter anbieten und deren Aufgabengebiete teils schwer messbar sind. Dies gilt für Schulen: Das Zentralabitur ist ein externer Eingriff in das Prüfungs- und Unterrichtsgeschehen, was in der bisherigen dezentralen Prüfungssituation in dieser Form nicht gegeben war. Bildung ist ein öffentliches, wenn nicht meritokratisches Gut, und schulische Qualitätsentwicklung ist nicht als einfache Produktionskette aufzufassen (Ditton, 2007, S. 88). Weiter ist die Anforderung, z.B. Alltagsbezüge im Unterricht herzustellen, um u.a. problembezogenes Denken zu trainieren<sup>1</sup>, erwünscht, aber komplex und kann daher nicht einfach durch zentrale Prüfungen gemessen und getestet werden.

---

1 Vgl. [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2008/2008\\_10\\_24-VB-EPA.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_10_24-VB-EPA.pdf) [21.01.2011].

### 2.3 Zentrale Prüfungen in Hessen und Bremen

In Hessen wie in Bremen wurden weitgehend inhaltliche Standards eingeführt („content standards“, Klieme, Avenarius, Blum, Döbrich, Gruber & Prenzel, 2003). Das heißt, die Art der Standards erfordert direkt eine inhaltliche Anpassung des Unterrichts an die zentralen Prüfungen. Allerdings ist eine bloße Ausrichtung des Unterrichts an den zentralen Prüfungen seitens des Ministeriums nicht erwünscht. Dies zeigen u.a. die zeitliche Begrenzung, die laut Lehrplan für das Lernen des Abiturstoffes vorgesehen ist, sowie die von der Regierung herausgegebenen, verpflichtenden Leitlinien für jedes Fach, die ebenfalls Themen vorgeben sowie auch pädagogische Ziele beinhalten.<sup>2</sup>

Weiter können die zentralen Prüfungen im internationalen Vergleich als low-stakes eingeordnet werden (Klein, Kühn, van Ackeren & Block, 2009), denn die Ergebnisse des Abiturs werden nicht veröffentlicht und es folgen auch keine Sanktionen. Allerdings haben sich die „stakes“ im innerdeutschen Vergleich über die Zeit durch die zentralen Prüfungen erhöht: zuerst existierten gar keine „stakes“, nun gibt es „low stakes“. Die Kontrolle wurde also ausgebaut und damit der Druck auf die Lehrenden erhöht.

Schließlich haben die Regierungen in Hessen und Bremen Unterstützungsmaßnahmen eingeführt. So werden den Lehrenden die Themen des Abiturs zwei bis zweieinhalb Jahre im Voraus mitgeteilt, Operatorenlisten und Übungsaufgaben werden herausgegeben. Die Regierungen in Hessen wie in Bremen greifen demnach kontrollierend *und* unterstützend in den Bildungsprozess ein. Dies könnte wiederum den Druck auf die Lehrpersonen abschwächen und entlastend sein.

## 3 Fragestellung und Hypothesen

In diesem Artikel interessiert die Frage, inwiefern es Lehrpersonen in einem zentralen Low-stakes-Prüfungssystem gelingt, Themenvarianz im Unterricht zu verwirklichen und Kongruenz zwischen Unterricht und zentraler Prüfung herzustellen. Themenvarianz wird hier u.a. verstanden als das Ausmaß, in dem die Interessen der Schüler/-innen und aktuelle Themen im Sachfach bei der inhaltlichen Auswahl der Unterrichtsthemen berücksichtigt werden. Sie kann in einem zentralen wie in einem dezentralen Prüfungssystem zu einem anregenden Unterricht beitragen. In diesem Artikel wird demnach die Themenvarianz in den zentral geprüften wie in den dezentral geprüften Fächern untersucht.

---

2 Vgl. [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2008/2008\\_10\\_24-VB-EPA.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_10_24-VB-EPA.pdf) [21.01.2011].

Kongruenz zwischen Unterricht und Prüfung ist in einem dezentralen Prüfungssystem keine Schwierigkeit, in einem zentralen jedoch eine Herausforderung. Hier werden daher die zentral geprüften Fächer fokussiert.

Es wird in diesem Beitrag angenommen, dass die beiden Konstrukte Themenvarianz und Kongruenz unter zentralen Prüfungsbedingungen voneinander unabhängig sind. Denn die Unwissenheit über die zentrale Prüfung selbst bringt es mit sich, dass die Strategie, Themen breit zu variieren ebenso wie Themen stark zu fokussieren, erfolgreich sein kann. Unter der Bedingung dezentraler Prüfungen wird dagegen angenommen, dass die Dimensionen nicht trennscharf sind, denn Kongruenz kann die einzelne Lehrperson unter dezentralen Bedingungen herstellen, unabhängig davon ob sie die Themen variiert oder nicht. Hier ist der Grad an Themenvarianz im Unterricht eher eine Frage des Unterrichtstils und keine Strategie, um die Schüler/-innen – hoffentlich – erfolgreich durch das Zentralabitur zu führen.

### 3.1 Themenvarianz

#### 1. *Zeichnet sich ein Teaching-to-the-test-Effekt ab und verändert sich dieser in Abhängigkeit der Erfahrung mit dem Zentralabitur?*

Hypothese T1a: Es wird angenommen, dass trotz der Unterstützungsmaßnahmen der Regierung und des Low-stakes-Prüfungssystems ein ttt-Effekt oder Verdrängungseffekt sowie der Motivation-Transfer-Effekt in Hessen und Bremen auftreten (Frey & Jegen, 2002). Denn das Instrument Zentralabitur ist im Vergleich zum Status quo kontrollierend und greift in die Autonomie der Lehrenden ein, zudem führt es durch die Einführung messbarer Indikatoren eine Hierarchie der Arbeitsbereiche ein (oberste Priorität haben die Inhalte des Abiturs). Dies bedeutet, dass zwischen den Lehrpersonen, die Kurse unterrichten, die zentral geprüft werden und denen, die Kurse unterrichten, die sie selbst prüfen (dezentrales Prüfungssystem), signifikante Unterschiede zu beobachten sein sollten.

Hypothese T1b: Es wird erwartet, dass der ttt-Effekt durch Erfahrung mit dem Zentralabitur entsteht und über die Jahre nicht wieder auf das Ursprungsniveau sinkt. So wird angenommen, dass der Verdrängungseffekt (Frey & Jegen, 2002) durch die Einführung des Zentralabiturs auftritt. Dies bedeutet, dass die Lehrpersonen die zentralen Prüfungen zu ihrem Kernarbeitsbereich zählen und Aspekte, von denen angenommen wird, dass sie nicht in den Prüfungen direkt getestet werden, vernachlässigt werden. Allerdings ist es fraglich, wie sich dieser Effekt über die Jahre – besonders langfristig – entwickelt. Dies ist abhängig davon, welche Erfahrungen die Lehrpersonen mit den von ihnen unterrichteten, aber zentral geprüften Fächern machen: Erreichen ihre Schüler/-innen das Abitur ebenso erfolgreich wie im dezentralen Prüfungssystem? Welche Art des Unter-

richtens war in 2007, 2008, 2009, ... erfolgreich; eine möglichst breite Interpretation der vorgegebenen Schwerpunktthemen oder aber eine starke Fokussierung? Entspricht sich diese Erfahrung zwischen den Jahren oder ist es jedes Jahr ein Glücksspiel, welche Art des Unterrichtens für die Schüler/-innen Erfolg versprechend ist? Da aus der jetzigen Perspektive und ohne eine detaillierte Analyse der Vorgaben der Schwerpunktthemen sowie der Abituraufgaben nicht abzusehen ist, wie die Erfahrung der Lehrpersonen langfristig aussehen wird, wird hier nur angenommen, dass in den ersten drei Jahren nach der Implementation der ttt-Effekt zu beobachten sein wird und sich nicht durch die Erfahrung abschwächt. Analysen nach fünf oder nach zehn Jahren nach der Implementation sowie fachspezifische Auswertungen müssten zeigen, ob die Erfahrung mit den zentralen Prüfungen die Lehrpersonen lehrt, dass sie den Unterricht besser stärker eingrenzen (ttt-Effekt würde weiter steigen) oder aber eher thematisch breit anlegen (ttt-Effekt würde zurückgehen) sollten. Denn die zentralen Prüfungen sind jedes Jahr neu und müssen auf lange Sicht variieren, wenn auch der Grad der Variation zwischen den Fächern (z.B. Mathematik gegenüber Englisch) unterschiedlich stark ausfallen kann.

2. *Welche Faktoren beeinflussen die Verwirklichung von Themenvarianz im Unterricht der zentral geprüften Fächer?*

Hypothese T2: Es wird angenommen, dass Lehrpersonen, die zentral geprüfte Fächer unterrichten, an Schulen, an denen das Kollegium sich als selbstwirksam wahrnimmt, vor einer Überkompensierung stärker geschützt sind als Lehrpersonen in einem Kollegium, welches sich nicht als selbstwirksam erlebt. Denn die Wahrnehmung des Kollektivs als selbstwirksam (Schwarzer & Jerusalem, 1999) kann ein Hinweis darauf sein, dass das Instrument Zentralabitur die Autonomie dieser Lehrpersonen nicht erschüttert hat. Ebenso wird davon ausgegangen, dass Lehrpersonen, die zu curricularen Themen kooperieren, die Unterrichtsinhalte variieren und z.B. Interessen der Schüler/-innen oder Tagesaktualitäten, die inhaltlich zum Fach passen, berücksichtigen. Kooperation in diesem Sinne wäre hilfreich für die Implementation der Gesetze (Seashore Louis, Marks & Kruse, 1996). Zudem wird erwartet, dass unsichere Lehrpersonen die Themen mehr einschränken als Lehrer/-innen, die sich nicht verunsichert fühlen (Liermann, 2009, S. 246). Schließlich wird angenommen, dass die Erfahrung mit den zentralen Prüfungen nicht dazu beiträgt, den ttt-Effekt zu minimieren (vgl. auch T1b). Als mögliche Drittvariable werden die Zugehörigkeit zu einem Bundesland und das Geschlecht der Lehrpersonen kontrolliert, für beide wird jedoch keine Hypothese aufgestellt.

### 3.2 Kongruenz zwischen Unterricht und zentraler Prüfung

#### 1. *Wie beurteilen die Lehrpersonen nach dem Abitur die Übereinstimmung zwischen der Vorbereitung im Unterricht und den zentralen Prüfungen?*

Hypothese K1: Es wird angenommen, dass es zwar möglich ist, in einem zentralen Prüfungssystem Kongruenz herzustellen, allerdings dürfte dies nicht allen Lehrpersonen gleichermaßen gelingen. Denn durch das Zentralabitur wurde die Rolle der Prüfenden und Lehrenden aufgebrochen und die Lehrenden sind nun unwissender gegenüber den gestellten Abituraufgaben im Vergleich zum dezentralen Abitur.

#### 2. *Verbessert sich durch die Erfahrung mit den zentralen Prüfungen die Kongruenz zwischen Unterricht und Prüfung aus der Perspektive der Lehrpersonen?*

Hypothese K2: Hier wird die Hypothese aufgestellt, dass sich aufgrund der eigenen Struktur zentraler Prüfungen im Schulsystem (wechselnde Themen und Fragestellungen im Abitur) und trotz „feedback loop“ die Kongruenz nicht oder nur begrenzt durch die Erfahrung mit den zentralen Prüfungen steigern lässt. Dies kann zudem zwischen den geprüften Fächern variieren. Diese Hypothese widerspricht der NPM-Theorie, denn nach dieser wird angenommen, dass eine erhöhte Kontrolle zu einer Anpassung zwischen Angebot (Unterricht) und Nachfrage (Inhalte des Abiturs) führe.

#### 3. *Gibt es Faktoren, die die Wahrnehmung der Kongruenz zwischen Unterricht und zentralen Prüfungen aus Sicht der Lehrpersonen erhöhen bzw. konterkarieren?*

Hypothese K3: In diesem Artikel wird angenommen, dass die Wahrnehmung kollektiver Selbstwirksamkeit (Schwarzer & Jerusalem, 1999), Kooperation zum Curriculum (Seashore Louis, Marks & Kruse, 1996) und unterstützende Maßnahmen seitens der Regierung (vgl. z.B. Frey & Jegen, 2002) für die Verwirklichung von Kongruenz hilfreich sind. Zu den unterstützenden Maßnahmen können beispielsweise Angaben zu den geprüften Themen, Operatoren und angebotene Fortbildungen gezählt werden. Von der Unsicherheit gegenüber den zentralen Prüfungen wird erwartet, dass sie in einem negativen Zusammenhang zur Kongruenz steht. Die Zugehörigkeit zu einem Bundesland und das Geschlecht der Lehrpersonen fungieren als Kontrollvariablen. Für sie werden keine Hypothesen aufgestellt.



## 4 Design und Methoden

In diesem Abschnitt werden die Stichprobe sowie die Variablen beschrieben. Zudem wird der Zusammenhang zwischen den untersuchten Konstrukten kurz dargestellt. Schließlich werden die Auswertungsverfahren dargelegt.

### 4.1 Stichprobe

Die Stichprobe ist für die Beantwortung der Fragen zur Themenvarianz und zur Kongruenz unterschiedlich, da die Fragen zur Themenvarianz in der standardisierten Befragung vor dem Abitur, die Fragen zur Kongruenz in jener nach dem Abitur gestellt wurden.

Die *Fragen zur Themenvarianz* haben Lehrpersonen der Klassen 12 und 13, die Kurse mit zentral und/oder dezentral geprüften Kursen unterrichteten, beantwortet. Dies ermöglicht es, Antworten zu zentral und dezentral geprüften Kursen direkt vergleichen zu können. Da jedoch nur in Bremen Kurse dezentral geprüft werden, sind nur im Bundesland Bremen Aussagen zu dezentral und zentral geprüften Kursen vorzufinden.

Insgesamt liegen aus allen drei Jahren 3917 Aussagen zur Dimension „Themenvarianz“ vor. 3273 Antworten beziehen sich dabei auf zentral geprüfte und 644 Antworten auf dezentral geprüfte Kurse. 355 der Lehrpersonen haben jedoch *innerhalb eines Jahres* dezentral wie zentral geprüfte Kurse unterrichtet, von ihnen liegen daher in diesem Jahr zwei Aussagen vor. Im Jahr 2007 liegen 174 zweifache Aussagen vor, im Jahr 2008 Aussagen von 95 Lehrpersonen und im Jahr 2009 von 86 Lehrpersonen. Dies reduziert die Stichprobengröße hinsichtlich der befragten Personen. Insgesamt können somit die Antworten von 3562 Personen ausgewertet werden. Tabelle 1 zeigt die Aufteilung der vorliegenden Aussagen auf die Jahre 2007, 2008 und 2009 sowie für Hessen und Bremen, inklusive der zweifachen Antworten.

*Tabelle 1: Anzahl der Aussagen, die zu den Variablen Themenvarianz in zentral und in dezentral geprüften Fächern vorliegen*

zentral geprüftes Fach	Themenvarianz		
	2007	2008	2009
N Bremen	395	394	319
N Hessen	766	681	718
dezentral geprüftes Fach			
N Bremen	309	183	152
N Hessen	-	-	-
Summe	1470	1258	1189

N = Anzahl Fälle

Betrachtet man nur die vorliegenden Aussagen zu den Fächern mit zentralen Prüfungen hinsichtlich der Themenvarianz, so zeigt sich, dass 44% der Aussagen (1445) von Lehrpersonen stammen, die nur einmal einen Kurs unterrichtet haben, der dann zentral geprüft wurde, 43% der Lehrpersonen haben 2mal (1384) und 13% 3mal Kurse unterrichtet, die zentral geprüft wurden.

Analysiert man die Aussagen zu den dezentral geprüften Kursen, so zeigt sich, dass 68% (439) der Aussagen von Lehrpersonen sind, die einmal in den drei untersuchten Jahren Kurse unterrichtet haben, die sie selbst prüften, 26% der Lehrpersonen haben 2mal (168) und lediglich 6% (37) der Lehrpersonen haben 3mal dezentral geprüfte Fächer unterrichtet.

Eine Besonderheit der Variable Themenvarianz ist, wie bereits bemerkt, dass für einen Teil der Lehrpersonen (355 Personen, 9%) *innerhalb eines Jahres zwei Antworten* zum Umgang mit der Themenvarianz (nämlich für die dezentral und zentral geprüften Kurse) vorliegen. Dies führt bei den mehrbenenanalytischen Auswertungen und den t-Tests, bei denen Aussagen zu dezentral und zentral geprüften Fächern ausgewertet werden, zu der Frage, welche der Aussagen berücksichtigt werden sollen. Denn beide Aussagen aufzunehmen, ist u.a. nicht sinnvoll, da ansonsten in den z.B. Mehrebenenanalysen auch alle Aussagen dieser Personen zu anderen Variablen doppelt im Datensatz vorhanden wären. Aus diesem Grund ist besonders für die mehrbenenanalytischen Auswertungen zu entscheiden, welche der Aussagen der Lehrpersonen in Bremen berücksichtigt werden, jene zu den zentral oder jene zu den dezentral geprüften Kursen. Um diese Frage zu klären, wurde vor den Analysen überprüft, ob sich die Aussagen der Lehrpersonen, die zentral *und* dezentral geprüfte Fächer innerhalb eines Schuljahres unterrichtet haben, von denen jener Lehrpersonen unterscheiden, die in einem Schuljahr *nur* zentral *oder* *nur* dezentral geprüfte Fächer anboten. Würden sich die Aussagen nicht signifikant voneinander unterscheiden, so könnte man frei wählen, ob die Aussagen zu den zentral oder dezentral geprüften Kursen verwendet würden. Unterscheidet sich eine Gruppe jedoch signifikant, so muss überlegt werden, wie damit im Weiteren umgegangen wird.

Tabelle 2 zeigt, dass sich die Aussagen zur Themenvarianz in den zentral geprüften Fächern nicht signifikant voneinander unterscheiden – unabhängig davon, ob eine Lehrperson zentrale *und* dezentral geprüfte Kurse in einem Jahr unterrichtet hat oder aber einzig zentral geprüfte. Hinsichtlich der Aussagen zu dezentral geprüften Fächern zeigen sich jedoch signifikante Unterschiede. Lehrpersonen, die dezentral geprüfte Fächer mit zentral geprüften direkt vergleichen konnten, kamen in 2007 und 2009 eher zu dem Schluss, dass sie in den dezentral geführten Fächern mehr Themenvarianz verwirklichten als Lehrpersonen, die nur dezentrale Kurse unterrichteten. Hierbei handelt es sich um mittlere Effekte (2007:  $d = 0.30$ , 2009:  $d = 0.34$ ). Es wurde daher entschieden, von den Lehrper-

sonen, von denen innerhalb eines Jahres Aussagen sowohl zu den dezentral wie auch zu den zentral geprüften Kursen vorliegen, die Antworten zu den zentralen Prüfungen in die Mehrebenenanalyse bzw. den t-Tests aufzunehmen. Denn so wird sichergestellt, dass die Einschätzungen zum dezentralen Abitur nicht überschätzt werden. Dies entspricht einer strengeren Überprüfung der Hypothese (dezentrale Prüfungen erlauben mehr Themenvarianz), als ein umgekehrtes Vorgehen.

*Tabelle 2: Vergleich der Aussagen zur Themenvarianz zwischen Lehrpersonen die zentrale und dezentrale Kurse unterrichtet haben und den Lehrpersonen, die entweder dezentrale oder zentrale Kurse unterrichtet haben*

	Themenvarianz			
zentral geprüfte Kurse	2007	2008	2009	Gruppe*
M	2.32 2.35	2.30 2.29	2.43 2.34	zentral beides
SD	.57 .65	.59 .68	.58 .65	zentral beides
N	987 174	980 95	951 86	zentral beides
t-Test	n.s.	n.s.	n.s.	
dezentral gepr. Kurse	2007	2008	2009	Gruppe*
M	2.86 3.00	2.93 2.93	2.77 2.93	dezentral beides
SD	.45 .49	.37 .52	.39 .53	dezentral beides
N	135 174	88 95	66 86	dezentral beides
t-Test	t (307) = 2.62 p < .01; d = 0.30	n.s.	t (150) = 2.15 p < .05; d = 0.34	

M = Mittelwert; SD = Standardabweichung; N = Anzahl Fälle; \*Gruppe = Gruppe der Lehrenden mit einer Aussage zu zentral *oder* dezentral geprüften Kursen bzw. der Lehrpersonen, von denen zwei Aussagen vorliegen, einmal zu ihrem Kurs, der dezentral und einmal zu ihrem Kurs, der zentral geprüft wurde; d = Cohens d Effektstärke; n.s. = nicht signifikant

Hinsichtlich der Frage nach der Kongruenz zwischen Unterricht und zentralen Prüfungen wurden einzig die Lehrer/-innen in der Nacherhebung befragt, die mit der Erstkorrektur eines zentral geprüften Kurses beauftragt waren. Es sind dies für Bremen 170 Lehrpersonen 2007, 215 Lehrpersonen 2008 und 176 Lehrpersonen im Jahr 2009. In Hessen liegen Daten von 388 Lehrpersonen 2007, 322 Lehrpersonen 2008 und 337 Lehrpersonen 2009 vor. Von diesen insgesamt 1608 Personen haben 1013 einmal (63%), 497 Personen 2mal (31%) und 98 Personen (6%) zu allen drei Erhebungszeitpunkten zentrale Prüfungen korrigiert.

## 4.2 Messinstrumente

Folgend werden zuerst die abhängigen Variablen, dann die unabhängigen Variablen (alphabetisch sortiert) anhand von Beispielitems beschrieben. Zudem werden zu den Variablen die Konsistenzwerte und Quellen angegeben.

- Themenvarianz (Vorerhebung; Skala; 5 Items; Cronbachs Alpha Hessen: Kurse mit zentralen Prüfungen:  $\alpha_{2007} = .75$ ,  $\alpha_{2008} = .77$ ,  $\alpha_{2009} = .76$ ; Bremen: Kurse mit zentralen Prüfungen:  $\alpha_{2007} = .77$ ,  $\alpha_{2008} = .80$ ,  $\alpha_{2009} = .81$ ; Bremen: Kurse mit dezentralen Prüfungen:  $\alpha_{2007} = .65$ ,  $\alpha_{2008} = .63$ ,  $\alpha_{2009} = .66$ ). Beispielitems: „Bei der Auswahl der Themen/Inhalte habe ich mir Zeit für die Bearbeitung von aktuellen Themen im Sachfach genommen.“ „Ich habe bei der Auswahl der Themen/Inhalte die Wünsche und Interessen der Schüler/-innen berücksichtigt.“ „Ich habe auch Themen gewählt, die nicht zum Prüfungsgebiet des Zentralabiturs gehören.“ Antwortformat: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu. Die Lehrpersonen in Hessen haben den Fragebogen für ihre zentral geprüften Kurse ausgefüllt, in Bremen konnten sie Angaben zu dezentral und zentral geprüften Kursen machen. Die Aussagen wurden in diesem Artikel für einzelne Analysen zu einer Variablen zusammengeführt, (vgl. auch Absatz zur Stichprobe sowie zu den Auswertungsverfahren; Kapitel 4.1 und 4.3 in diesem Beitrag). Quelle: Eigenentwicklung.
- Kongruenz Lehrpersonen (Nacherhebung; Skala; 4 Items; Cronbachs Alpha Hessen:  $\alpha_{2007} = .77$ ,  $\alpha_{2008} = .78$ ,  $\alpha_{2009} = .74$ ; Bremen:  $\alpha_{2007} = .80$ ,  $\alpha_{2008} = .81$ ,  $\alpha_{2009} = .73$ ). Beispielitems: „Zwischen den in den Abiturprüfungen berücksichtigten und den in meinem Unterricht bearbeiteten Inhalten gibt es eine hohe Übereinstimmung.“ „Zwischen den Abiturprüfungen und meinen Klausuren gibt es eine hohe Kongruenz in Bezug auf das gestellte Anforderungsniveau.“ „Zwischen den Abiturprüfungen und meinen Klausuren gibt es eine hohe Kongruenz in den erforderlichen Bearbeitungsprozessen.“ Antwortformat: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu. Quelle: Eigenentwicklung.
- Beteiligung am Zentralabitur (BeteiligungZA): Mindestens ein Kurs mit zentraler Prüfung wird im jeweiligen Jahr unterrichtet: 1 = ja; 0 = nein.
- Bundesland: 1 = Bremen; 0 = Hessen.
- Erfahrung mit dem Zentralabitur (ErfahrungZA): 0 = gar keine Erfahrung, 1 = ein Jahr Erfahrung, 2 = zwei Jahre Erfahrung; 3 = drei Jahre Erfahrung.
- Fortbildung (Nacherhebung; Einzelitem): „Hilfreiche Grundlagen für die angemessene Vorbereitung auf die zentralen Abiturprüfungen waren die Fortbildungsveranstaltungen zu den Schwerpunktthemen.“ Antwortformat: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft ge-

nau zu. Quelle: Hamburger Landesinstitut für Lehrerbildung und Schulentwicklung (2006).

- Geschlecht: 1 = weiblich, 0 = männlich.
- Jahr: 0 = 2007, 1 = 2008, 2 = 2009.
- Kollektive Selbstwirksamkeit (Koll. Selbstwirksamkeit) (Vorerhebung; Skala; 5 Items; Cronbachs Alpha Hessen:  $\alpha_{2007} = .67$ ,  $\alpha_{2008} = .77$ ,  $\alpha_{2009} = .75$ ; Bremen:  $\alpha_{2007} = .69$ ,  $\alpha_{2008} = .73$ ,  $\alpha_{2009} = .72$ ). Beispielitem: „Bitte geben Sie an, inwieweit die folgenden Aussagen in Ihrer Schule zutreffen. Auch mit außergewöhnlichen Vorfällen können wir zurechtkommen, da wir uns im Kollegium gegenseitig Rückhalt bieten.“ Antwortformat: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu. Quelle: Schwarzer & Jerusalem (1999).
- Kooperation hinsichtlich des Curriculums (Kooperation Curriculum) (Vorerhebung; Skala; 5 Items; Cronbachs Alpha Hessen:  $\alpha_{2007} = .90$ ,  $\alpha_{2008} = .89$ ,  $\alpha_{2009} = .90$ ; Bremen:  $\alpha_{2007} = .89$ ,  $\alpha_{2008} = .90$ ,  $\alpha_{2009} = .87$ ). Beispielitems: „Entscheidung über die Reihenfolge der Unterrichtsthemen“, „Vorbereitung des Unterrichts bzw. der Einheiten und Kurse“. Antwortformat: 5 = einmal wöchentlich, 4 = einmal monatlich, 3 = mehrmals im Halbjahr, 2 = einmal im Jahr, 1 = gar nicht. Quelle: Steinert, Gerecht, Klieme & Döbrich (2003).
- Operatoren (Nacherhebung; Einzelitem): „Hilfreiche Grundlagen für die angemessene Vorbereitung auf die zentralen Abiturprüfungen war die Liste der Operatoren.“ Antwortformat: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu. Quelle: Eigenentwicklung.
- Themen (Nacherhebung; Einzelitem): „Hilfreiche Grundlagen für die angemessene Vorbereitung auf die zentralen Abiturprüfungen waren die Angaben zu den verbindlichen Schwerpunktthemen für das Abitur.“ Antwortformat: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu. Quelle: Hamburger Landesinstitut für Lehrerbildung und Schulentwicklung (2006).
- Unsicherheit gegenüber dem Zentralabitur (UnsicherheitZA) (Vorerhebung; Skala; 4 Items; Cronbachs Alpha Hessen:  $\alpha_{2007} = .70$ ,  $\alpha_{2008} = .69$ ,  $\alpha_{2009} = .71$ ; Bremen:  $\alpha_{2007} = .71$ ,  $\alpha_{2008} = .74$ ,  $\alpha_{2009} = .71$ ). Beispielitem: „Ich habe Angst, dass ein Thema kommt, in dem die Schüler/-innen nicht gut vorbereitet sind.“ Antwortformat: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu. Quelle: Eigenentwicklung.

### 4.3 Auswertungsverfahren

Die Daten der befragten Lehrpersonen wurden deskriptiv ausgewertet, Differenzen wurden mittels eines t-Tests untersucht. Zudem wurden die Daten mittels des Softwareprogramms „Hierarchical Linear and Nonlinear Modeling“ (HLM 6.06) von Raudenbush, Bryk und Congdon (2004) analysiert.

Bei den mehrebenenanalytischen Verfahren zur Themenvarianz sowie zur Übereinstimmung zwischen Unterricht und Abiturprüfung wurden mittels HLM Dreiebenenmodelle berechnet. Auf der ersten Ebene sind alle Variablen verortet, die über die Zeit variieren. Dies sind hinsichtlich der Analyse zur Themenvarianz, bei der die Antworten zu dezentral und zu zentral geprüften Fächern gemeinsam ausgewertet wurden (A): Das Jahr, die Beteiligung am Zentralabitur im jeweiligen Jahr und die Erfahrung mit dem Zentralabitur. Bei der Analyse zur Themenvarianz, bei der nur die Antworten zu zentral geprüften Kursen analysiert wurden (B), wurden das Jahr, die Erfahrung mit dem Zentralabitur, die Angaben zur kollektiven Selbstwirksamkeit, zur Unsicherheit gegenüber den zentralen Prüfungen und zur Kooperation hinsichtlich des Curriculums auf der ersten Ebene aufgenommen. Hinsichtlich der Frage zur Kongruenz zwischen Unterricht und zentraler Prüfung wurden auf der ersten Ebene folgende Variablen berücksichtigt: Das Jahr, die Erfahrung mit dem Zentralabitur, Angaben zur kollektiven Selbstwirksamkeit, zur Unsicherheit gegenüber den zentralen Prüfungen und zur Kooperation hinsichtlich des Curriculums, zudem Aussagen zu Unterstützungsmaßnahmen (Angaben zu den Schwerpunktthemen, zu Fortbildungen und zu der Operatorenliste). Fehlende Werte sind bei Mehrebenenanalysen auf der ersten Ebene kein Problem, da eine gleiche Anzahl von Beobachtungen nicht vorausgesetzt wird (Hox, 2002, S. 95f).

Auf der zweiten Ebene befinden sich die Variablen, die zwischen den Individuen, hier Lehrpersonen, unterschiedlich, aber über die Zeit konstant sind. Hier wurde bei allen Analysen das Geschlecht aufgenommen. Die dritte Ebene schließlich bezieht die Differenzen zwischen den Schulen mit in die Analyse ein, hier wurde das Bundesland berücksichtigt.

Die folgenden Regressionsgleichungen wurden eingesetzt. Dabei sind die Variablen kursiv gedruckt, die zentriert (grand mean) in das Modell aufgenommen wurden:

- Themenvarianz  $A^3 = \gamma_{000} + \gamma_{001} * \text{Bundesland} + \gamma_{010} * \text{Geschlecht} + \gamma_{100} * \text{BeteiligungZA} + \gamma_{200} * \text{ErfahrungZA} + \gamma_{300} * \text{Jahr} + r_0 + u_{00} + e$

3 Bei der Analyse A zur Themenvarianz wurden die Antworten zu dezentral und zu zentral geprüften Fächern gemeinsam ausgewertet.

- Themenvarianz  $B^4 = \gamma_{000} + \gamma_{001} * \text{Bundesland} + \gamma_{010} * \text{Geschlecht} + \gamma_{100} * \text{BeteiligungZA} + \gamma_{200} * \text{ErfahrungZA} + \gamma_{300} * \text{Jahr} + \gamma_{400} * \text{Koll. Selbstwirk.} + \gamma_{500} * \text{UnsicherheitZA} + \gamma_{600} * \text{Kooperation Curriculum} + r_0 + u_{00} + e$
- Kongruenz  $= \gamma_{000} + \gamma_{001} * \text{Bundesland} + \gamma_{010} * \text{Geschlecht} + \gamma_{100} * \text{ErfahrungZA} + \gamma_{200} * \text{Jahr} + \gamma_{300} * \text{UnsicherheitZA} + \gamma_{400} * \text{Kooperation Curriculum} + \gamma_{500} * \text{Koll. Selbstwirk.} + \gamma_{600} * \text{Themen} + \gamma_{700} * \text{Fortbildung} + \gamma_{800} * \text{Operatoren} + r_0 + u_{00} + e$

## 5 Ergebnisse Themenvarianz

Tabelle 3 zeigt, dass in den dezentral geprüften Fächern die Lehrpersonen eher der Aussage zustimmen, Themenvarianz umsetzen zu können. So geben Lehrpersonen aus Bremen, die dezentrale Fächer unterrichten, an, Themen zu variieren (2007, 2008 und 2009  $M > 2.50$ ). Dagegen wird die Themenvarianz nach Aussagen der Lehrpersonen in den Fächern mit zentralen Prüfungen in Bremen sowie in Hessen eher nicht realisiert (2007, 2008, 2009  $M < 2.50$ ). Leider lässt sich in Hessen nicht zurückverfolgen, wie das Niveau bezüglich der Themenvarianz im dezentralen Prüfungssystem beschaffen war.

In allen drei Jahren unterscheiden sich zudem *innerhalb von Bremen* die Aussagen der Lehrenden zur Umsetzung der Themenvarianz signifikant zwischen zentral und dezentral geprüften Fächern (Ergebnisse des t-Tests, vgl. Tabelle 3). Es handelt sich dabei um mittlere bis große Effekte (Cohens  $d = 0.85$ ,  $d = 0.92$  und  $d = 0.60$ ).

Der *Vergleich der Bremer Ergebnisse mit den hessischen zu den zentralen Prüfungen* zeigt in den Jahren 2007 und 2008 signifikante Unterschiede. Die hessischen Lehrpersonen geben an, die Themenvarianz stärker einzuschränken als ihre Bremer Kolleg/-innen. Die Effektstärke (Cohens  $d = -0.16$ ,  $d = -0.27$ ) ist allerdings klein. In 2009 nähern sich die Werte der beiden Bundesländer an und es zeigt sich kein signifikanter Unterschied mehr (vgl. Tabelle 3).

Die mehrbenenanalytischen Ergebnisse des Längsschnitts bestätigen die deskriptiven Befunde (vgl. Tabelle 4): So zeigt das Gesamtmodell, dass die Beteiligung am Zentralabitur ( $p < .001$ ) im jeweiligen Jahr mit einer Eingrenzung der Themenvarianz zusammenhängt. Das heißt, Lehrpersonen, die an zentralen Prüfungen beteiligt sind, grenzen die Themen mehr ein als Lehrpersonen, die dezentrale Kurse unterrichten. Für die Variable „Jahr“ ( $p < .01$ ) und das Bundesland Bremen ( $p < .01$ ) ist ein signifikanter Zuwachs an Themenvarianz zu beobachten. Das bedeutet, dass die Themenvarianz über die drei Jahre mehr zunimmt. Zudem setzen die Lehrpersonen in Bremen nach eigenen Aussagen mehr

4 Bei der Analyse B zu Themenvarianz wurden nur die Antworten zu zentral geprüften Kursen analysiert.

Themenvarianz um als ihre Hessener Kolleg/-innen. Die Variable „Erfahrung mit den zentralen Prüfungen“ und das Geschlecht werden nicht signifikant.

*Tabelle 3: Lehrpersonen – Deskriptive Auswertungen – Themenvarianz*

	Themenvarianz		
Bremen zentral	2007	2008	2009
M	2.38	2.40	2.41
SD	.60	.61	.63
N	395	394	319
Bremen dezentral	2007	2008	2009
M	2.94	2.93	2.86
SD	.48	.45	.48
N	309	183	152
zentral-dezentral*	t (308) = -9.70 p < .001; d = 0.85; n <sub>z</sub> = 395; n <sub>dz</sub> = 135	t (209) = -10.64 p < .001; d = 0.92; n <sub>z</sub> = 394; n <sub>dz</sub> = 88	t (145) = -5.92 p < .001; d = 0.60; n <sub>z</sub> = 319; n <sub>dz</sub> = 66
	Themenvarianz		
Hessen zentral	2007	2008	2009
M	2.29	2.24	2.42
SD	.57	.58	.57
N	766	681	718
Bremen: zentral – Hessen: zentral*	t (1159) = 2.55 p < .05; d = -0.16; n <sub>Bz</sub> = 395; n <sub>H<sub>z</sub></sub> 766	t (1073) = 4.18 p < .001; d = -0.27; n <sub>Bz</sub> = 394; n <sub>H<sub>z</sub></sub> = 681	n.s.

M = Mittelwert; SD = Standardabweichung; N = Anzahl Fälle; z = zentral geprüft; dz = dezentral geprüft; Bz = Bremen zentral geprüfte Fächer; Hz = Hessen zentral geprüfte Fächer; Antwortformat Themenvarianz: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu.

\*Bremen: Die deskriptiven Auswertungen zu zentralen und dezentralen Prüfungen pro Jahr beinhalten auch die Lehrpersonen mit zweifachen Antworten. Beim Vergleich (t-Test) Bremen-zentral versus Bremen-dezentral und Bremen-zentral versus Hessen-zentral wurden jedoch nur die Antworten von den Lehrpersonen in Bremen mit zweifachen Antworten zu den *zentralen* Prüfungen berücksichtigt (vgl. Stichprobe, Kapitel 4.1 in diesem Beitrag).

Die Standardisierung der Variablen weist darauf hin, dass die Beteiligung an zentralen Prüfungen (-0.21) unter den berücksichtigten Faktoren den stärksten Prädiktor darstellt. Die Zugehörigkeit zu einem Bundesland (0.09) und der Einfluss der Jahre (0.05) stehen demgegenüber in einem schwächeren Zusammenhang.

Die mehrbenenanalytischen Ergebnisse aus Tabelle 4 zeigen zudem, dass von der Gesamtvarianz rund 44% durch die Veränderung über die Zeit innerhalb der Individuen und rund 52% durch die Unterschiede zwischen den Lehrpersonen erklärt werden können. Differenzen zwischen den Schulen erklären 4% der Varianz.



*Tabelle 4: Ergebnisse der Längsschnittanalysen für die abhängige Variable Themenvarianz unter Berücksichtigung der Antworten zu dezentral und zentral geprüften Fächern (A)*

Fixe Effekte	Modell 1: Intercept-Only	Modell 2: Gesamtmodell
<b>Ebene 1</b>		
Konstante, $B_0$	2.41 (0.02)***	2.77 (0.05)***
BeteiligungZA (Teilnahme am ZA = 1)		-0.47 (0.04)***
ErfahrungZA		n.s.
Jahr		0.04 (0.01)**
<b>Ebene 2</b>		
Geschlecht (weiblich = 1)		n.s.
<b>Ebene 3</b>		
Bundesland (Bremen = 1)		0.10 (0.04)**
<b>Zufällige Effekte</b>		
Varianzkomponenten		
E3: $u_{00}(\tau_{p0})$	0.01***	0.01***
E2: $r_0(\tau_{a0})$	0.19***	0.18***
E1: $c(\sigma^2)$	0.16	0.15
Devianz	5522	5355
N der geschätzten Parameter	4	9

Schätzer: unstandardisierte Koeffizienten (Standardfehler); E = Ebene; +  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; n.s. = nicht signifikant; Modell 1 und Modell 2: N Ebene 1 = 3431, N Ebene 2 = 1929, N Ebene 3 = 37; Antwortformat Themenvarianz: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu; ZA = Zentralabitur

In Tabelle 5 sind die Ergebnisse zu den Einflussfaktoren auf die Themenvarianz hinsichtlich zentral geprüfter Fächer dargestellt. Ähnlich den mehrbenenanalytischen Auswertungen zu der Themenvarianz zu dezentral und zentral geprüften Fächern werden von der Gesamtvarianz rund 43% durch die Veränderung über die Zeit, rund 55% durch die Unterschiede zwischen den Lehrpersonen und 2% der Varianz durch die Differenzen zwischen den Schulen erklärt.

Das Gesamtmodell in Tabelle 5 zeigt, dass die Unsicherheit gegenüber den zentralen Prüfungen ( $p < .01$ ) mit einer Eingrenzung der Themenvarianz zusammenhängt. Das heißt, Lehrpersonen, die gegenüber den zentralen Prüfungen angeben, unsicher zu sein, grenzen die Themen mehr ein als Lehrpersonen, die nicht angeben, wegen des Zentralabiturs unsicher zu sein.

Lehrpersonen, die in zentral geprüften Fächern die Themen variieren, nehmen dagegen eher eine kollektive Selbstwirksamkeit wahr ( $p < .001$ ), kooperieren eher zum Curriculum ( $p < .05$ ) und unterrichten eher in Bremen als in Hessen ( $p < .01$ ). Für die Variable „Jahr“ ist zudem ein signifikanter Zuwachs an Themenvarianz zu beobachten ( $p < .01$ ). Die Variable „Erfahrung mit den zentralen Prüfungen“ und das Geschlecht werden nicht signifikant.

*Tabelle 5: Ergebnisse der Längsschnittanalysen für die abhängige Variable Themenvarianz unter Berücksichtigung der Antworten einzig zu den zentral geprüften Fächern (B)*

Fixe Effekte	Modell 1: Intercept-Only	Modell 2: Gesamtmodell
<b>Ebene 1</b>		
Konstante, $B_0$	2.36 (0.02)***	2.32 (0.03)***
ErfahrungZA		n.s.
Jahr		0.04 (0.01)**
Koll. Selbstwirksamkeit		0.17 (0.03)***
UnsicherheitZA		-0.09 (0.03)**
Kooperation Curriculum		0.03 (0.01)*
<b>Ebene 2</b>		
Geschlecht (weiblich = 1)		n.s.
<b>Ebene 3</b>		
Bundesland (Bremen = 1)		0.10 (0.04)**
<b>Zufällige Effekte</b>		
Varianzkomponenten		
E3: $u_{00}(\tau_{\beta 0})$	0.01***	0.01**
E2: $\tau_0(\tau_{\pi 0})$	0.19***	0.18***
E1: $e(\sigma^2)$	0.15	0.15
Devianz	5054	4679
N der geschätzten Parameter	4	11

Schätzer: unstandardisierte Koeffizienten (Standardfehler); E = Ebene; +  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; n.s. = nicht signifikant; Modell 1: N Ebene 1 = 3172, N Ebene 2 = 1783, N Ebene 3 = 37; Modell 2: N Ebene 1 = 3024, N Ebene 2 = 1718, N Ebene 3 = 37; Antwortformat Themenvarianz: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu; ZA = Zentralabitur

Standardisiert man die Variablen, so zeigt sich, dass die Wahrnehmung der kollektiven Selbstwirksamkeit (standardisierter Koeffizient = 0.20) unter den untersuchten Variablen am engsten mit der Themenvarianz im Zusammenhang steht. Mit Abstand folgt die Zugehörigkeit zum Bundesland Bremen (0.12) und der negative Einfluss der Unsicherheit gegenüber den zentralen Prüfungen (-0.09). Der Zusammenhang zwischen Themenvarianz und Jahr (0.03) bzw. Themenvarianz und der Kooperation zum Curriculum (0.02) ist dagegen eher marginal.

## 6 Ergebnisse Kongruenz

Die deskriptiven Befunde in Tabelle 6 zeigen, dass generell in Hessen wie in Bremen in allen drei Jahren der Aussage eher zugestimmt wird, dass die Vorbereitung im Unterricht mit den geprüften Inhalten und dem Anforderungsniveau im Abitur übereinstimmt ( $M < 2.5$ ).

Tabelle 6: Lehrpersonen – Deskriptive Auswertungen pro Jahr – Kongruenz

	Kongruenz		
Bremen	2007	2008	2009
M	3.09	3.08	3.09
SD	0.49	0.52	0.48
N	170	215	176
Hessen	2007	2008	2009
M	3.08	3.05	3.02
SD	0.48	0.47	0.48
N	388	322	337

M = Mittelwert; SD = Standardabweichung; N = Anzahl Fälle; Antwortformat Kongruenz: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu

Analysiert man die Fragestellung, welche Faktoren für die Wahrnehmung der Kongruenz zwischen Abiturprüfungen und Unterricht relevant sind, so zeigt Tabelle 7 die längsschnittlichen Ergebnisse der Mehrebenenanalyse.

Der größte Anteil der Gesamtvarianz wird durch die intraindividuellen Unterschiede erklärt (61%). 38% der Differenzen in der Kongruenz zwischen Abiturprüfung und Unterricht lassen sich auf Varianzen zwischen den Individuen zurückführen. Nur ein marginaler Anteil, 1%, kann durch den Faktor Schule erklärt werden.

Tabelle 7 zeigt, dass eine hohe Übereinstimmung zwischen Unterricht und Abiturprüfung mit der Einschätzung einhergeht, dass die Schwerpunktthemen ( $p < .001$ ), Fortbildungen ( $p < .05$ ) und Operatoren ( $p < .05$ ) unterstützend sind. Ebenfalls steht die Wahrnehmung kollektiver Selbstwirksamkeit in einem positiven Zusammenhang mit der Kongruenz zwischen Unterricht und Prüfung, allerdings nur auf dem 10%-Niveau. Hessische Lehrpersonen nehmen ihren Unterricht etwas stärker als kongruent zu den zentralen Abiturprüfungen wahr als Bremer Lehrpersonen ( $p < .05$ ). Unsicherheit gegenüber den zentralen Prüfungen ( $p < .01$ ) und die Erfahrung mit den zentralen Prüfungen ( $p < .10$ ) sind dagegen hinderlich für die Verwirklichung von Kongruenz. Faktoren wie das Jahr, die Kooperation zum Curriculum und das Geschlecht stehen in keinem Zusammenhang mit der erzielten Kongruenz.

Die Standardisierung der Koeffizienten zeigt, dass die Vorgabe zu den Schwerpunktthemen (standardisierter Koeffizient = 0.24) und (mit Abstand) die Unsicherheit gegenüber den zentralen Prüfungen (-0.11) am stärksten im Zusammenhang mit der Herstellung von Kongruenz zwischen Unterricht und Abiturprüfung stehen. Es folgen die Erfahrung mit den zentralen Prüfungssystem (-0.08), die Fortbildungen (0.08), die Zugehörigkeit zu einem Bundesland (-0.08), die Angaben zu den Operatoren (0.07) und die kollektiven Selbstwirksamkeitserwartungen (0.05).

Tabelle 7: Lehrpersonen – Kongruenz zwischen Unterricht und den Prüfungen im Zentralabitur (Längsschnitt)

Fixe Effekte	Modell 1: Intercept-Only	Modell 2: Gesamtmodell
<b>Ebene 1</b>		
Konstante, $B_0$	3.07 (0.02)***	3.09 (0.02)***
Jahr		n.s.
ErfahrungZA		-0.04 (0.03) <sup>+</sup>
UnsicherheitZA		-0.09 (0.03)**
Kooperation Curriculum		n.s.
Koll. Selbstwirksamkeit		0.05 (0.03) <sup>+</sup>
Themen		0.16 (0.02)***
Fortbildung		0.04 (0.02)*
Operatoren		0.04 (0.02)*
<b>Ebene 2</b>		
Geschlecht (weiblich = 1)		n.s.
<b>Ebene 3</b>		
Bundesland (Bremen = 1)		-0.07 (0.03)*
<b>Zufällige Effekte</b>		
Varianzkomponenten		
E3: $u_{00}(\tau_{\beta 0})$	0.00	0.00
E2: $r_0(\tau_{\pi 0})$	0.09***	0.07***
E1: $e(\sigma^2)$	0.14	0.11
Devianz	1990	1239
N der geschätzten Parameter	4	14

Schätzer: unstandardisierte Koeffizienten (Standardfehler); E = Ebene; <sup>+</sup>  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; n.s. = nicht signifikant; Modell 1: N Ebene 1 = 1524, N Ebene 2 = 1036, N Ebene 3 = 37; Modell 2: N Ebene 1 = 1113, N Ebene 2 = 793, N Ebene 3 = 37; Antwortformat Kongruenz: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu

## 7 Diskussion

### 7.1 Themenvarianz

Die deskriptiven Befunde sowie die mehrbenenanalytischen Auswertungen zeigen, dass auch unter Low-stakes-Bedingungen Themenvarianz im zentralen Prüfungssystem eher vernachlässigt wird als im dezentralen Prüfungssystem, also ein ttt-Effekt bzw. Verdrängungseffekt auftritt (Hypothese T1a). In Bezug auf den ttt-Effekt muss berücksichtigt werden, dass die Gruppe der Lehrenden im dezentralen Prüfungssystem klein und nur in Bremen vorzufinden ist. Die durchgeführten t-Tests zeigen zum einen, dass sich innerhalb von Bremen zentrale und dezentrale Kurse hinsichtlich der Themenvarianz signifikant unterscheiden. Zum anderen wurde durch die t-Tests zwischen den Aussagen zu den zentral geprüften Fächern in Hessen und Bremen überprüft, ob die Ergebnisse in Hessen zu

den zentralen Prüfungen denen aus Bremen zu den zentralen Prüfungen ähneln. Hier konnte beobachtet werden, dass die hessischen Lehrer/-innen in 2007 und 2008 signifikant häufiger angaben, die Themenvarianz einzuschränken als ihre Bremer Kolleg/-innen, wobei es sich um mittlere Effekte handelt.

Wie können diese Unterschiede erklärt werden? Zum einen könnte die Themenvarianz in Hessen bereits vor der Einführung der zentralen Tests begrenzt gewesen sein. Zum anderen könnte die Einrichtung der zentralen Prüfungen zeitgleich in allen Fächern für die stärkere Fokussierung verantwortlich sein. Interessanterweise gleichen sich die Ergebnisse in 2009 zwischen Hessen und Bremen an, so dass kein signifikanter Unterschied mehr zu beobachten ist. Der signifikante Unterschied zwischen dezentralen und zentralen Prüfungen bleibt in Bremen auch 2009 erhalten. Allerdings sinkt die Effektstärke von 2007 bis 2009. Da parallel dazu die Mittelwerte zu den zentral geprüften Kursen nahezu konstant bleiben, wäre es interessant zu analysieren, ob in den folgenden Jahren die Themenvarianz auch in den dezentral geprüften Kursen weiter sinkt, wie es die Annahme eines Motivation-Transfer-Effekt wäre. Die Prüfungskultur der zentralen Fächer würde dann auf die der dezentral geprüften Fächer übertragen werden. Frey und Jegen (2002, S. 38) weisen auf die Problematik hin, dass es nicht absehbar ist, was in Bereichen geschieht, in denen (wie in den dezentral geprüften Fächern) die intrinsische Motivation vorausgesetzt wird, aber durch die Einführung extrinsischer Anreize in anderen Bereichen unterhöhlt wird. In diesen Bereichen kann durch den Motivation-Transfer-Effekt die intrinsische Motivation reduziert werden. Im Gegensatz zu den zentral geprüften Fächern wird jedoch nicht versucht, die intrinsische Motivation durch extrinsische Motivation zu ersetzen. Es entsteht also ein Vakuum und es ist nicht klar, wie die Lehrpersonen darauf reagieren.

Hinsichtlich der Interpretation der deskriptiven Befunde in Bremen und Hessen könnte aber auch vermutet werden, dass sich die Entwicklung in Hessen (Sinken der Themeneingrenzung) fortsetzt und vielleicht auch in Bremen einsetzt. Allerdings scheint eine solche Entwicklung aufgrund der Werte in Bremen zu den zentralen Prüfungen (in den untersuchten drei Jahren nahezu konstant) als eher unwahrscheinlich. In diesem Beitrag wird eher vermutet, dass die Überkompensierung in Hessen aufgrund der einheitlichen Einführung zuerst besonders stark war, sich dann jedoch auf einem Niveau einpendeln wird, welches dem in Bremen ähnelt.

Auch die Hypothese T1b, dass die Themenvarianz durch die Erfahrung mit den zentralen Prüfungen nicht tangiert wird, bestätigt sich. Dies, so die eingangs formulierte These, liegt daran, dass die Einschränkung der Themenvarianz durch die Verdrängung der intrinsischen Motivation durch externe Anreize auftritt. Dies kann nicht durch Erfahrung mit den zentralen Prüfungen, also Erfahrung

mit externen Anreizen, vollständig kompensiert werden. Es wäre interessant zu untersuchen, wie sich dieser Effekt langfristig (z.B. fünf oder zehn Jahre nach der Implementation) entwickelt. Denn es ist denkbar, dass sich der Effekt verschärft oder abschwächt, je nach dem, welche Erfahrung die Lehrpersonen mit dem Prüfungssystem machen. Schwächt die Erfahrung die Unsicherheit ab oder verstärkt sie sie?

Die Hypothesen (T2) zu den förderlichen bzw. hinderlichen Faktoren für Themenvarianz in den zentral geprüften Fächern können weitgehend angenommen werden: Hohe kollektive Selbstwirksamkeitsüberzeugungen, niedrige Unsicherheit und Kooperation zum Curriculum stehen in Zusammenhang mit einem Unterricht, in dem Themen variiert und die Interessen der Schüler/-innen berücksichtigt werden. Allerdings sind die Effekte gering. Zudem nimmt die Themenvarianz über die Jahre zu und wird von den Bremer Lehrpersonen eher umgesetzt als von den hessischen Lehrpersonen. Für das Geschlecht und die Erfahrung mit den zentralen Prüfungen ist kein Zusammenhang zur Themenvarianz zu beobachten.

## 7.2 Kongruenz

Die Ergebnisse dieser Analysen zeigen, dass es den Lehrenden eher gelingt, Kongruenz zwischen Unterricht und zentraler Prüfung herzustellen (Hypothese K1). Hinsichtlich der Kongruenz konnte leider kein Vergleich zwischen dezentral und zentral geprüften Fächern durchgeführt werden. Allerdings kann angenommen werden, dass Lehrpersonen im dezentralen Abitur mehrheitlich den Items zur Kongruenz nahezu voll und ganz zustimmen würden, da sie die Fragestellungen individuell entwickeln. Unter dieser Annahme lässt sich festhalten, dass es so scheint, als sei es für die Lehrpersonen im zentralen Prüfungssystem eine Herausforderung geworden, Kongruenz herzustellen.

In Hypothese K2 wurde formuliert, dass der Unterricht durch die Erfahrung mit den zentralen Prüfungen an die Abiturprüfungen nicht besser angepasst werden kann. Auch diese Hypothese bestätigt sich. Das Ergebnis zu dem Einfluss der Erfahrung verschärft sogar noch die hier vertretene These: Auf dem 10%-Niveau wird ein negativer Zusammenhang sichtbar. Dies bedeutet, dass je mehr Erfahrungen jemand mit den zentralen Prüfungen hat, desto weniger gelingt es, Kongruenz herzustellen (vgl. Tabelle 7). Dieser negative Zusammenhang sollte in zukünftigen Studien näher betrachtet werden. Hier wird damit die durch das New Public Management versprochene Steigerung der Steuerungsfähigkeit der Regierung über Zeit und Erfahrung (Schedler, 1995) nicht bestätigt.

Schließlich wurde in diesem Artikel nach Einflussfaktoren für eine gelungene Herstellung von Kongruenz zwischen Unterricht und Abiturprüfung ge-

fragt. Die Analyse bestätigt weitgehend die eingangs formulierte Hypothese (K3): So ist ein geringes Unsicherheitsempfinden sowie die Wahrnehmung des Kollegiums als kollektiv selbstwirksam förderlich für die Herstellung der Kongruenz zwischen Unterricht und Prüfung. Die unterstützenden Maßnahmen (Angaben von Themen, Fortbildungen und die Operatorenliste) stehen ebenfalls in einem positiven Zusammenhang mit der Umsetzung von Kongruenz. Hinsichtlich des Bundeslandes zeigte sich, dass es den Lehrpersonen in Hessen nach eigenen Aussagen eher gelingt, Kongruenz herzustellen als den Lehrpersonen in Bremen. Die Kooperation zum Curriculum, das Merkmal Geschlecht und das Jahr wurden nicht signifikant.

Dass der Faktor „Kooperation zum Curriculum“ in keinem signifikant positiven Zusammenhang mit der Verwirklichung von Kongruenz steht, widerspricht der formulierten Hypothese. Ursache dafür kann sein, dass Kooperation zwar unterstützend sein kann, bei Unwissenheit aber auch Unsicherheit verstärkt, z.B. in Teams mit vielen unsicheren Lehrpersonen. Weitere Untersuchungen müssten zeigen, welche Ursachen im Detail für dieses Ergebnis verantwortlich sind.

Die Vorgaben zu den Schwerpunktthemen stehen mit Abstand am engsten in Zusammenhang mit der Herstellung von Kongruenz. Hier muss allerdings angemerkt werden, dass die Einschätzungen zu den Unterstützungsmaßnahmen erst nach dem Abitur erhoben wurden und somit nachdem die Lehrpersonen wussten, ob sie ihre Schüler/-innen erfolgreich durch das Abitur geführt haben oder nicht. Es ist demnach unklar, inwiefern die Maßnahmen *vor* dem Abitur den Lehrpersonen halfen, Unsicherheit abzubauen. Die Ergebnisse zeigen aber immerhin, dass eine signifikante Mehrheit Kongruenz herstellen konnte und diesen Erfolg auch auf die vom Ministerium bereitgestellten Informationen zurückführt.

Betrachtet man nun abschließend die Analysen zum Thema Themenvarianz in den zentral geprüften Fächern (Analyse B) und Kongruenz, so zeigen die Mehrebenenanalysen – wie erwartet –, dass beide Konstrukte durch unterschiedliche Faktoren beeinflusst werden. So nimmt die Themenvarianz über die Jahre zu (kleiner Effekt), die Kongruenz zwischen Unterricht und Prüfung kann jedoch über die Jahre nicht signifikant gesteigert werden. Die Themenvarianz wird in Bremen leichter erreicht, in Hessen dagegen wird die Kongruenz zwischen Unterricht und Prüfung nach Aussagen der Lehrpersonen eher umgesetzt. Kooperation zum Curriculum steht in einem positiven Zusammenhang zur Themenvarianz (kleiner Effekt), hat aber keinen Effekt auf die Herstellung von Kongruenz zwischen Unterricht und Abiturprüfung. Die Erfahrung mit zentralen Prüfungen steht mit der Themenvarianz in keinem Zusammenhang, hat jedoch hinsichtlich der Kongruenz einen leichten negativen Effekt.

Neben den Unterschieden zeigen sich jedoch auch Gemeinsamkeiten: Kollektive Selbstwirksamkeitserwartungen sind mit beiden Konstrukten positiv kor-

reliert, wenn auch der Effekt hinsichtlich der Themenvarianz deutlich stärker ist. Ebenso steht die Unsicherheit gegenüber den zentralen Prüfungen zu beiden Konstrukten in einem negativen Verhältnis. Schließlich ist das Merkmal Geschlecht bei beiden Konstrukten nicht signifikant.

### 7.3 *Fazit*

Abschließend lässt sich festhalten, dass das Steuerungsinstrument Zentralabitur auch in einem Low-stakes-Prüfungssystem und trotz Unterstützungsmaßnahmen der Regierungen im Zusammenhang mit einer über das Erfordernis hinausgehenden Eingrenzung der Themenvarianz im Unterricht steht. Die Ursachen, so die Annahme hier, liegen darin, dass die eingeführten Standards weitgehend „content standards“ sind (Klieme et al., 2003), die am Ende der Schulzeit geprüft werden. Das Zentralabitur stellt zudem im innerdeutschen, historischen Vergleich eine Erhöhung der „stakes“ dar. Dies wurde im Sinne der Steigerung der Steuerungsfähigkeit seitens der Ministerien angestrebt. Das heißt, die Autonomie der Lehrenden wurde eingeschränkt und eine Hierarchie zwischen den Arbeitsaufträgen (Vorgaben zum Abitur – Curriculum) wurde eingeführt.

Die Lehrenden erzielen – aufgrund der Unterstützungsmaßnahmen und trotz Unsicherheit gegenüber den zentralen Prüfungen – eine Anpassung ihres Unterrichts an die abgefragten Inhalte im Abitur. Allerdings zeigt sich durch die Erfahrung mit den zentralen Prüfungen keine Steigerung der Kongruenz zwischen Unterricht und Abiturprüfung. Stattdessen ist sogar ein leicht negativer Effekt zu beobachten. Dies entspricht der Hypothese, dass das NPM-Konzept hier nicht auf das Bildungssystem übertragen werden kann. Abituraufgaben sollen jedes Jahr neu und unbekannt sein. Lehrpersonen können demnach den Unterricht nicht durch einen „feedback loop“ besser an die Prüfungen anpassen. So kann, je nachdem, welche Erfahrungen die Lehrpersonen sammeln, der ttt-Effekt weiter ansteigen oder aber auch teilweise abnehmen. Hier wird die (weiter zu überprüfende) Annahme getroffen, dass die Themenvarianz aufgrund einer „Restunsicherheit“ gegenüber den zentralen Prüfungen jedoch nicht wieder so variiert wird wie in einem dezentralen Prüfungssystem.

Im Weiteren sind Untersuchungen interessant, die den ttt-Effekt fachspezifisch und zudem langfristig in Bremen und Hessen und/oder in Staaten/Bundesländern analysieren, die schon seit geraumer Zeit zentrale Prüfungen eingeführt haben. Auch wäre es interessant, Schüler/-innendaten zu analysieren und weitere Aspekte des Unterrichtens neben der Themenvarianz in die Analyse miteinzubeziehen. Schließlich wären Analysen der Schwerpunktthemen im Vergleich zu den Abituraufgaben sinnvoll, um abschätzen zu können, welche Erfahrungen die Lehrpersonen mit den zentralen Prüfungen sammeln: Zeigen die Abi-



turaufgaben, dass eine Engführung der Schwerpunktthemen oder aber eine breite Interpretation hilfreich für die Vorbereitung gewesen wären? Das heißt, können die Lehrpersonen langfristig Strategien im Umgang mit den zentralen Prüfungen aufbauen, die dann auch erfolgreich sind, oder verändern sich die Anforderungen Jahr für Jahr?

## 8 Literatur

- Abrams, L.M. & Madaus, G.F. (2003). The lessons of high-stakes testing. *Educational Leadership*, 61 (3), 31-35.
- Abrams, L.M. (2007). Implications of high-stakes testing for the use of formative classroom assessment. In J.H. McMillan (Ed.), *Formative classroom assessment. Theory into practice* (pp. 79-98). New York/London: Teacher College, Columbia University.
- Au, W. (2007). High-stakes testing and curricular control: a qualitative metasynthesis. *Educational Researcher*, 36 (5), 258-267.
- Bishop, J.H. (1995). The impact of curriculum-based external examinations on school priorities and student learning. *International Journal of Educational Research*, 23 (8), 653-752.
- Bogumil, J. & Jann, W. (2009). *Verwaltung und Verwaltungswissenschaft in Deutschland: Einführung in die Verwaltungswissenschaft* (2., völlig überarb. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brozo, W.G. & Hargis, C. (2003). Using low-stakes reading assessment. *Educational Leadership*, 61 (3), 60-64.
- Ditton, H. (2007). Schulqualität – Modelle zwischen Konstruktion, empirischen Befunden und Implementierung. In J. van Buer & C. Wagner (Hrsg.), *Qualität von Schule. Ein kritisches Handbuch* (S. 83-92). Bern: Peter Lang Verlag.
- Frey, B. & Jegen, R. (2002). Kontraproduktive Wirkung des Motivators 'Geld': Intrinsische Motivation. *Verbands-Management*, 28 (3), 30-41.
- Hamburger Landesinstitut für Lehrerbildung und Schulenwicklung (2006). *Schriftliche Abiturprüfungen 2006*. Schülerfragebogen/Lehrerfragebogen. Hamburg.
- Hamilton, L.S. & Stecher, B.M. (2002). Improving test-based accountability. In L.S. Hamilton, B.M. Stecher & S.P. Klein (Eds.), *Making sense of test-based accountability in education* (pp. 121-144). Santa Monica, CA: RAND Education.
- Hamilton, L.S., Stecher, B.M., Russell, J.L., Marsh, J.A. & Miles, J. (2008). Accountability and teaching practices: school-level actions and teacher responses. In B. Fuller, M.K. Henne & E. Hannum (Eds.), *Strong stakes, weak schools: the benefits and dilemmas of centralized accountability* (pp. 31-66). Bingley: Emerald Group Publishing Limited.
- Hoppe, C. (2004). Die Fortbildungsoffensive. In Projektleitung "Selbständige Schule" (Hrsg.), *Verantwortung für Qualität – Grundlagen des Projektes* (Bd.1; S. 75-83). Troisdorf: Bildungsverlag EINS.
- Klein, E.D., Kühn, S.M., Ackeren, I. van & Block, R. (2009). Wie zentral sind zentrale Prüfungen? – Abschlussprüfungen am Ende der Sekundarstufe II im nationalen und internationalen Vergleich. *Zeitschrift für Pädagogik*, 55 (4), 596-621.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., et al. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Bonn: Bundesministerium für Bildung und Forschung.
- Liermann, C. (2009). Individualkonzepte unter dem Einfluss von Standardisierungsprozessen in der Schule. In J. Bilstein & J. Ecarius (Hrsg.), *Standardisierung - Kanonisierung. Erziehungswissenschaftliche Reflexionen* (S. 235-251). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Niskanen, W.A. (1971). *Bureaucracy and representative government*. Chicago: Aldine Atherton.
- Organisation for Economic Co-operation and Development (OECD). (2004). *Bildung auf einen Blick. OECD-Indikatoren 2004*. Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- Raudenbush, S.W., Bryk, A.S. & Congdon, R. (2004). *HLM 6 for Windows [Computer software]*. Lincolnwood, IL: Scientific Software International, Inc.
- Richter, I. (2004). Chancen und Risiken selbständiger Schulen in Nordrhein-Westfalen. In Projektleitung "Selbständige Schule" (Hrsg.), *Verantwortung für Qualität – Grundlagen des Projektes*, (Bd.1; S. 1-29). Troisdorf: Bildungsv Verlag E1ns.
- Schedler, K. (1995). Ansätze einer wirkungsorientierten Verwaltungsführung: Von der Idee des New Public Managements (NPM) zum konkreten Gestaltungsmodell; Fallstudie Schweiz. Bern: Paul Haupt.
- Schwarzer, R. & Jerusalem, M. (Hrsg.) (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin und Humboldt-Universität Berlin.
- Seashore Louis, K., Marks, H. & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal*, 33 (4), S. 757-798.
- Stecher, B.M. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In L.S. Hamilton, B.M. Stecher & S.P. Klein (Eds.), *Making sense of test-based accountability in education* (pp. 79-100). Santa Monica: Rand.
- Steffens, U. (2007). Schulqualitätsdiskussion in Deutschland. Ihre Entwicklung im Überblick. In J. van Buer & C. Wagner (Hrsg.), *Qualität von Schule. Ein kritisches Handbuch* (S. 21-51). Bern: Peter Lang Verlag.
- Steinert, B., Gerecht, M., Klieme, E. & Döbrich, P. (2003). *Skalen zur Schulqualität: Dokumentation der Erhebungsinstrumente. Arbeitsplatzuntersuchung (APU), Pädagogische Entwicklungsbilanzen (PEB)*. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung.
- Weiß, M. (2003). Kann das Schulwesen durch Wettbewerb genesen? In: H. Döbert, B. von Kopp, R. Martini & M. Weiß (Hrsg.), *Bildung vor neuen Herausforderungen, Historische Bezüge – rechtliche Aspekte – Steuerungsfragen - Internationale Perspektiven* (S. 111-119). Neuwied: Luchterhand.

Daniela J. Jäger, 2012: Schulklima, Selbstwirksamkeit und Arbeitszufriedenheit aus Sicht der Lehrpersonen und Schüler/-innen in Hessen und Bremen. In: Maag Merki, Katharina (Hrsg.): *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland*. Wiesbaden: Springer VS, S. 65-93.

Hinweis

The final publication is available at  
[http://link.springer.com/chapter/10.1007/978-3-531-94023-6\\_4](http://link.springer.com/chapter/10.1007/978-3-531-94023-6_4).

#### **4. Schulklima, Selbstwirksamkeit und Arbeitszufriedenheit aus Sicht der Lehrpersonen und Schüler/-innen in Hessen und Bremen**

*Daniela J. Jäger*

Schulklima, individuelle sowie kollektive Selbstwirksamkeitsüberzeugungen und Arbeitszufriedenheit sind wichtige Indikatoren für Schulqualität. So steht ein als angenehm wahrgenommenes Schulklima im Zusammenhang mit einer Reihe positiver psychosozialer Wirkungen wie beispielsweise einer günstigen Selbstkonzeptentwicklung bei den Schüler/-innen (Eder, 1994). Individuelle Selbstwirksamkeitsüberzeugungen, die Schwarzer und Schmitz (1999) definieren als die subjektive Gewissheit, eine neue oder schwierige Aufgabe auch bei Widerständen erfolgreich bearbeiten zu können, werden als relevant betrachtet, da sie den Umgang mit Herausforderungen für den Einzelnen erleichtern. Analog dazu kann angenommen werden, dass ein Kollektiv (wie ein Lehrer/-innenkollegium) ebenfalls Herausforderungen erfolgreicher bewältigen wird, wenn es überindividuell von der Handlungskompetenz der Gruppe überzeugt ist (Schwarzer & Jerusalem, 1999, S. 85). Arbeitszufriedenheit ist schließlich ein wichtiger Indikator für Lebenszufriedenheit allgemein, steht jedoch auch im Zusammenhang mit Leistung und Motivation (van Dick, 2006). Ist Arbeitszufriedenheit gegeben, so werden sich, laut Ergebnissen der Industriesoziologie, eine gesteigerte Konfliktlösekompetenz und eine höhere Produktivität der Individuen einstellen (zitiert nach Fend, 1998).

Auch die erfolgreiche Implementation von Reformvorhaben, wie die Einführung von zentralen Prüfungen, kann eher gelingen, wenn die Einzelnen das Schulklima als positiv wahrnehmen, von ihrer Wirksamkeit und der des Kollegiums überzeugt sind und zudem mit ihrer Arbeit zufrieden sind.

Daher werden nun in diesem Artikel die Faktoren Schulklima, Arbeitszufriedenheit, individuelle und kollektive Selbstwirksamkeit an den Schulen in Bremen und Hessen untersucht. Besonders interessant ist, ob sich die Einführung des Zentralabiturs auf diese Aspekte schulischen Lebens auswirkt. Die folgenden drei Fragestellungen sollen beantwortet werden:

1. Wie werden die Aspekte Schulklima, Arbeitszufriedenheit und individuelle sowie kollektive Selbstwirksamkeit von den Lehrpersonen über die Jahre wahrgenommen? Wie nehmen Schüler/-innen das Schulklima wahr?

2. Zeigt sich ein Zusammenhang zwischen der Erfahrung der Lehrpersonen mit dem zentralen Prüfungssystem und den genannten Konstrukten?
3. Stehen persönliche und biografische Faktoren im Zusammenhang mit den untersuchten Konstrukten?

Als Datengrundlage werden die Antworten der Lehrpersonen und der Schüler/-innen aus den Jahren 2007, 2008 und 2009 genutzt.

## **1 Begriffsdefinitionen und kurzer Literaturüberblick**

Zunächst werden die Konzepte Schulklima, Arbeitszufriedenheit und individuelle sowie kollektive Selbstwirksamkeit definiert und diskutiert. Zusätzlich wird kurz auf die Ursachen sowie auf die Effekte der Faktoren eingegangen.

### *1.1 Schulklima*

Die Schulvergleichsforschung stellte in den 1970ern/1980ern fest, dass sich Unterschiede zwischen den Schulen nicht erschöpfend durch Rahmenbedingungen, Budget, Schulform, Curriculum und beispielsweise rechtliche Bestimmungen erklären lassen (Buholzer, 2000, S. 25). Denn Schulen, die unter scheinbar den gleichen Bedingungen arbeiteten, erzielten unterschiedliche Ergebnisse hinsichtlich der pädagogischen Qualität und der Leistungen der Schüler/-innen. Fend (1998, S. 48) fasst daher rückblickend zusammen, dass der Begriff des Schulklimas erst einmal „Prozessmerkmale von Bildungssystemen beschreiben sollte, die auf kulturelle Gestaltungskräfte zurückgehen“. Auf diese Weise sollten fehlende Erklärungsfaktoren identifiziert und untersucht werden.

Seitdem werden Begriffe wie Schulkultur, Schulklima, Klassen-, Unterrichts- und Organisationsklima debattiert (vgl. z.B. Bessoth & Weibel, 2003; Eder, 1994; Fend, 1977, 1986, 1998; Helsper, Böhme, Kramer & Lingkost, 2001; Rutter, Maughan, Mortimore & Ouston, 1980). Dabei hat sich bis dato kein einheitlicher oder feststehender Topos zu den Begrifflichkeiten herauskristallisiert und auch eine fehlende Theoriebildung wird kritisiert (Gruehn, 2000, zitiert nach Halbheer & Kunz, 2011). Die Abgrenzungen zwischen den Begrifflichkeiten scheinen darüber hinaus diffus (Buholzer, 2000, S. 26; Halbheer & Kunz, 2011, S. 129).

Das Schulklima ist Teil der Schulkultur. Diese definieren Bessoth und Weibel (2003, S. 22) als das Norm- und Wertsystem der Organisationsmitglieder insgesamt. Der Begriff Schulklima umfasst demnach die „praktizierte Kultur“

und berücksichtigt tatsächliche Verhaltensweisen und gelebte Normen (Bessoth & Weibel, 2003, S. 22).

In diesem Artikel wird das Schulklima zudem als ein personalistisches Konstrukt verstanden und weniger als ein Organisationsmerkmal (Eder, 1996, S. 78). Das bedeutet, dass das Schulklima als „psychologisches Klima“ verstanden wird, welches „im strengen Sinne lediglich die Alltagstheorien einer Person über Merkmale der sie umgebenden Umwelt abbildet“ (Eder, 1996, S. 78). Das Schulklima spiegelt somit „die phänomenologischen Überzeugungen der Betroffenen (Schüler, Eltern, Lehrpersonen) hinsichtlich der schulischen Umwelt“ (Eder, 1996, S. 57). Es ist also in erster Linie – wie Litwin und Stringer (1968, S. 43, zitiert nach Bessoth & Weibel, 2003, S. 27) festhalten – nicht wichtig, was in der Organisation geschieht, sondern wie das Geschehene wahrgenommen wird. Dem stimmt auch Fend (1998, S. 48) zu, wenn er hervorhebt, dass die Wahrnehmung der schulischen Umwelt keine einfache Widerspiegelung der Realität ist, sondern vielmehr „eine subjektive, aber häufig mit Kollegen validierte Konstruktion“, die inspiriert sei „von Erfahrungen in anderen Schulen, von Ansprüchen, wie etwas sein sollte, von persönlichen Erfolgserfahrungen oder Erniedrigungen“. Die Erfassung dieses psychologischen Klimas ist relevant, da das wahrgenommene Klima als ein Filter verstanden werden kann, durch den die objektiven Phänomene hindurch müssen (Litwin & Stringer, 1968, S. 43, zitiert nach Bessoth & Weibel, 2003, S. 27). Individuen handeln erst – so die Annahme – aufgrund der Interpretation ihrer Wahrnehmung und diese erfolgt subjektiv wie auch kollektiv.

Dies bedeutet, dass zur Erfassung des Schulklimas ein „multi-source“ Ansatz gewählt werden sollte (Fend, 1998, S. 48). Denn erst durch die Auswertung der Daten zu der Perzeption der Lehrpersonen, Schüler/-innen, Eltern und der Schulleitung kann im Idealfall das Gesamtklima der Schule erfasst werden. Dies kann dann als Indikator für Problemkonstellationen und Problemlösungskapazitäten der Schule dienen, wie Halbheer und Kunz (2011) im Rückgriff auf Fend (1998) festhalten.

Das Schulklima aus der Sicht der Lehrpersonen ist, laut Forschungsergebnissen, relativ invariant gegenüber persönlichen und biografischen Merkmalen, jedoch sensibel gegenüber Kontextmerkmalen der Schule (Thomas, 1976): So beispielsweise gegenüber der Lage der Schule. Studien, die die verschiedenen Schultypen (Grundschule, Hauptschule, Realschule, Gymnasium, Gesamtschule) berücksichtigen, stellten systematische Unterschiede fest (vgl. z.B. Ulich, 1998, S. 382, zitiert nach Varbelow, 2003). Buholzer (2000, S. 38) nennt als wichtige Einflussfaktoren den Schultyp und das Schulsystem, die Lage und Größe der Schule sowie Merkmale der Schüler/-innen und Lehrpersonen.

Forschungsergebnisse zu Effekten des Schulklimas auf eine Leistungssteigerung bei den Schüler/-innen ergeben keine eindeutigen Ergebnisse, hingegen

berichten verschiedene Autor/-innen von positiven psychosozialen Wirkungen (vgl. für einen aktuellen, kurzen Überblick Halbheer & Kunz, 2011, S. 128). So fanden beispielsweise Oswald, Pfeifer, Ritter-Berlach und Tanzer (1989, S. 199) in ihrer Studie heraus, dass ein als positiv wahrgenommenes Schulklima mit weniger Schulangst und Schulunlust zusammenhängt. Eders (1994, S. 413) Wirkungsanalysen haben gezeigt, „dass sich Verhaltens- und Befindensmerkmale (Mitarbeit im Unterricht, Schulzufriedenheit, Belastungen durch die Schule) außerordentlich gut aus dem Klima der Schule vorhersagen lassen; Leistungen und Persönlichkeitsmerkmale (Selbstkonzeptentwicklung, Interessen) sind ebenfalls signifikant“.

Schließlich untersuchte Varbelow (2003) den Zusammenhang zwischen abweichendem Verhalten und Schulklima und stellte fest, dass ein negatives soziales Klima auch eher mit abweichenden Verhaltensweisen (wie Störungen im Unterricht, aber auch delinquentes Verhalten) einhergeht. Der Einfluss des Schulklimas bezieht sich demnach vor allem auf psychosoziale Faktoren, wobei das Schulklima teilweise auch als Moderatorvariable hinsichtlich der Leistungsfähigkeit der Schüler/-innen angesehen wird (Eder, 1996).

## 1.2 Arbeitszufriedenheit

Für den Begriff der Arbeitszufriedenheit gibt es bislang keine eindeutige Begriffsbestimmung und trotz langjähriger Forschung keine umfassende Theorie, wie Ammann (2004, S. 15f.) resümiert. Als allgemeine Definition kann das Verständnis von Bruggemann, Groskurth und Ulich (1975, S. 13) dienen, welche die Arbeitszufriedenheit als „Zufriedenheit mit einem gegebenen betrieblichen Verhältnis“ (in Anlehnung an den englischen Begriff „job satisfaction“) verstehen.

Das Konstrukt der Arbeitszufriedenheit ist schwer zu fassen, weil es zum Ersten auf die Arbeitsstelle und deren Bedingungen, zum Zweiten auf die Tätigkeit selbst und zum Dritten auf die Person im Arbeitskontext verweist. Zudem hat sich in der Theorie sowie in der Empirie gezeigt, dass sich der Begriff der Arbeitszufriedenheit nicht trennscharf vom Konstrukt der Arbeitsmotivation unterscheiden lässt (Six & Kleinbeck, 1989, S. 379, zitiert nach van Dick, 2006, S. 111).

Spieß und Winterstein (1999, S. 127) heben beispielsweise die arbeitsrelevanten Faktoren hervor, wenn sie Arbeitszufriedenheit definieren als eine „relativ stabile, d.h. zeitlich eher langfristig gleich bleibende Wertung der arbeitsrelevanten Gegebenheiten“. Die „arbeitsrelevanten Gegebenheiten“ umfassen dabei eine Vielzahl arbeits- und organisationsbezogener Faktoren, wie beispielsweise den Arbeitsinhalt, die Arbeitsanforderungen, die Autonomie der Mitarbeitenden, das Arbeitsumfeld und die Gesundheit, soziale Beziehungen zu Kolleg/-innen

und Vorgesetzten, Lohn- und Sozialleistungen, Karrierechancen und die Unternehmenskultur.

Ergebnisse aus Umfragen in den 1970ern und 1980ern machen den subjektiven Aspekt des Konstruktes Arbeitszufriedenheit deutlich: So gaben in Umfragen 80% bis 90% der Befragten an, sie seien mit ihrer Arbeit zufrieden oder sehr zufrieden (für einen Überblick siehe Ammann, 2004, S. 71f. und S. 118; Grunder & Bieri, 1995, S. 63; van Dick, 2006, S. 71). Gleichzeitig wurde teilweise in denselben Studien jedoch die Belastung am Arbeitsplatz von den Befragten hoch eingeschätzt (Merz, 1979a, zitiert nach Ammann, 2004, S. 60). Wenn man die hohen Fehlzeiten und beispielsweise die Fluktuation der Mitarbeiter/-innen mitbedenkt, so schließen Grunder und Bieri (1995, S. 63), dass Zweifel an dem Sinn und an dem praktischen Nutzen des Konstruktes Arbeitszufriedenheit angebracht seien. Sie begründen ihre Ansicht, indem sie hervorheben, dass das Konstrukt der Arbeitszufriedenheit eher subjektive Sachverhalte bzw. Befindlichkeiten wiedergibt und von daher nicht auf die Bewertung der objektiven Qualität der Arbeitsbedingungen übertragen werden dürfe. In diesem Sinne interpretierte auch Bauer (1980, zitiert nach Grunder & Bieri, 1995, S. 77) hohe Zufriedenheitsbekenntnisse gerade für den Lehrberuf dahingehend, dass die befragten Lehrer/-innen keine tiefen Arbeitszufriedenheitswerte nennen, „weil manifestierte Berufsunzufriedenheit als Eingeständnis persönlichen Versagens gelten dürfe“. Dies liegt vor allen Dingen daran, dass der Lehrberuf in Deutschland Beamtenstatus genießt und die Arbeit der Lehrenden zumeist isoliert stattfindet. Eine ähnliche Erklärung bietet das Prozessmodell der Arbeitszufriedenheit von Bruggemann, Groskurth und Ulich (1975, S. 132). Sie betonen, dass Arbeitszufriedenheit subjektiv hergestellt wird und abhängig ist von dem Anspruchsniveau, der Problemwahrnehmung und der Situationsbewältigung.

Für die Entstehung und die Veränderung der Arbeitszufriedenheit sind demnach, so fassen Spieß und Winterstein (1999, S. 132) zusammen, situative sowie persönliche Faktoren verantwortlich. So fand Merz (1979b, 1980, zitiert nach van Dick, 2006, S. 126) heraus, dass Lehrerinnen zufriedener seien als Lehrer, Lehrpersonen von Grundschulen zufriedener als Lehrpersonen an Hauptschulen oder Gymnasien und dass sich die Zufriedenheit mit dem steigenden Dienstalter ebenfalls erhöhe. Ammann (2004, S. 77ff.) dagegen berichtet, dass hinsichtlich der Auswirkungen des Geschlechts auf die Arbeitszufriedenheit uneinheitliche Resultate vorliegen und auch die Ergebnisse zum Einfluss des Dienst- und Lebensalters streuen. So lässt sich teilweise ein U-förmiger und teilweise ein linearer Zusammenhang beobachten (je länger eine Lehrperson beschäftigt bzw. je älter die Lehrperson ist, desto zufriedener ist sie mit ihrer Arbeit). Schließlich heben Spieß und Rosenstiel (2010) sowie van Dick (2006) die



Bedeutung relativer Autonomie am Arbeitsplatz hervor, die die intrinsische Motivation steigern und zu mehr Arbeitszufriedenheit führen würde.

Als Effekte von Arbeitszufriedenheit finden sich in der Literatur u.a. die folgenden Ausführungen: So steht die Arbeitszufriedenheit in einem positiven Zusammenhang mit Lebenszufriedenheit bzw. besserem psychologischen Wohlbefinden und einer geringeren Ausprägung psychosomatischer Beschwerden (Ammann, 2004, S. 71; Bruggemann, Groskurth & Ulich, 1975, S. 138; Fend, 1998, S. 107; van Dick, 2006, S. 118). Van Dick (2006, S. 118) diskutiert die Beobachtung, dass Arbeitszufriedenheit und Leistung sowie Leistung und Motivation in einem Zusammenhang stehen. Er betont, dass die Richtung des Zusammenhangs allerdings bisher nicht herausgearbeitet werden konnte. Fend (1998, S. 104) referiert die Ergebnisse der industriesoziologischen Forschung dahingehend, dass mit einer geringen Arbeitszufriedenheit ein geringes subjektives Wohlbefinden einhergeht und die Konfliktlösungsfähigkeit in der sozialen Einheit sowie die Produktivität reduziert werden. Zudem würden aufgrund fehlender Arbeitszufriedenheit kleinste Anlässe zu sozialen Konflikten, Widerstand und Abwehrkämpfen führen (Fend, 1998, S. 104). Höhere Arbeitszufriedenheit dagegen, so wird angenommen, würde zu geringeren Fehlzeiten, weniger Fluktuation und weniger Absentismus führen (van Dick, 2006, S. 118; Spieß & Rosenstiel, 2010, S. 83).

Hinsichtlich der Arbeitszufriedenheit von Lehrpersonen hebt Fend (1998, S. 104) hervor, dass die Arbeitszufriedenheit ein wichtiger Indikator für die Schulqualität sei. Je größer die Arbeitszufriedenheit, desto geringer würden Lehrpersonen Gleichgültigkeit, Anonymität und Resignation berichten (Fend, 1998, S. 105f.). Die Arbeitszufriedenheit sei daher mitunter ein Anzeichen für ein gutes soziales Klima in einem Kollegium und könne die Intensität des sozialen Lebens der Lehrpersonen abbilden. Dies würde auch, laut Fend (1998, S. 106), ein besseres Vertrauensverhältnis der Schüler/-innen zum Lehrkollegium hervorrufen. Er betont aber, dass die Reichweite dieses Zusammenhangs aufgrund empirischer Ergebnisse als gering einzuschätzen ist. So konnte bisher keine Wirkung des verbesserten Vertrauensverhältnisses auf die Steigerung der Leistungsbereitschaft, des Selbstvertrauens oder der Disziplin der Schüler/-innen nachgewiesen werden (Fend, 1998, S. 106). Opdenakker und van Damme (2006, S. 8f.) finden allerdings schwache Zusammenhänge zwischen der Arbeitszufriedenheit und der Fähigkeit der Lehrperson, die Klasse zu führen sowie zwischen einem schüler/-innenzentrierten Unterricht und der Arbeitszufriedenheit der Lehrperson.

### 1.3 Individuelle und kollektive Selbstwirksamkeit

Das Konzept der Selbstwirksamkeitserwartung basiert auf der sozial-kognitiven Theorie Banduras (1993) und beinhaltet das Vertrauen einer Person, eine Situation mittels eigener Kompetenzen und eigener Anstrengung zu meistern – auch wenn sich Schwierigkeiten in den Weg stellen.

Der Grad der Selbstwirksamkeitserwartung, die auch als Kompetenzerwartung bezeichnet wird, beeinflusst nach Schmitz und Schwarzer (2000) die Auswahl von Handlungen (Schwierigkeitsgrad), die investierte Anstrengung im Zielerreichungsprozess und die Ausdauer angesichts von Schwierigkeiten. Indirekt wirken sich demnach Kompetenzerwartungen auch auf den Grad des Handlungserfolgs aus. Denn je selbstwirksamer sich eine Person empfindet, desto schwierigere Aufgaben traut sie sich zu und desto ausdauernder wird sie an der Lösung einer Aufgabe arbeiten als eine Person, die sich als weniger selbstwirksam erlebt. Die Chance, letztlich auch erfolgreich zu sein, steigt – zumindest bei gleichen Kompetenzen. Hohe Selbstwirksamkeitsüberzeugungen unter den Lehrpersonen können dazu beitragen, dass Lehrer/-innen auch in neuen Situationen mit Schüler/-innen und Eltern oder der Schulleitung selbstsicherer interagieren als Lehrpersonen mit niedrigen Selbstwirksamkeitsüberzeugungen.

Bandura (1993, 1997) unterscheidet die Selbstwirksamkeitserwartung anhand zweier Dimensionen: der Generalitätsdimension und der Dimension, welche individuelle und kollektive Erwartungen beinhaltet. Die Generalitätsdimension umfasst folgende drei Aspekte: die *allgemeine Selbstwirksamkeitsüberzeugung*, welche sich auf alle Lebensbereiche bezieht, die *bereichsspezifische Selbstwirksamkeitsüberzeugung* (z.B. auf den Arbeitsbereich) und die *situationsspezifische*. Letztere besagt, dass die Selbstwirksamkeitserwartung je nach der spezifischen Situation (z.B. überraschend versus alltäglich) variieren kann.

Die Unterscheidung zwischen individueller und kollektiver Kompetenzerwartung auf der zweiten Dimension hat Bandura erst 1997 seinen Ausführungen zur Selbstwirksamkeit hinzugefügt. Schwarzer und Schmitz (1999) weisen daher darauf hin, dass es sich bei der kollektiven Selbstwirksamkeit noch um ein junges Konstrukt handelt, zu dem es bisher keine sozialpsychologische Grundlagen-, sondern lediglich Anwendungsforschung gibt. Sie zeigen jedoch gleichzeitig in ihrer Studie, dass sich die Konstrukte individuelle und kollektive Selbstwirksamkeit auch empirisch unterscheiden. So ist die individuelle Selbstwirksamkeit zwar der wichtigste Prädiktor für kollektive Selbstwirksamkeitserwartungen, erklärt aber dennoch nur 25% der Varianz des Konstruktes (Schwarzer & Schmitz, 1999). Bandura (1997, S. 477f.) hatte dies bereits theoretisch erörtert, indem er festhielt, dass die kollektive Selbstwirksamkeitserwartung mehr ist als

die Summe der individuellen Erwartungen. Dies liegt vor allem daran, dass Gruppen vor andere Herausforderungen gestellt werden als Individuen (z.B. Gruppenkohäsion), Ziele geteilt werden müssen und das Handeln der Mitglieder koordiniert werden muss. Von der kollektiven Selbstwirksamkeit werden demnach analoge positive Effekte erwartet wie von der individuellen Selbstwirksamkeit, sie beruht jedoch auf anderen Voraussetzungen, was wiederum auch variierende Effekte bedingen kann. Dies zeigt beispielsweise ein Ergebnis von Schmitz (2001, S. 64): Sie fand heraus, dass bei Frauen eine wahrgenommene kollektive Kompetenzerwartung ein wichtiger Schutzfaktor vor Burnout darstellt, bei Männern jedoch nicht.

Bei der Interpretation von Selbstwirksamkeitserwartungen muss berücksichtigt werden, dass beim Aufbau der Kompetenzerwartung nicht allein die tatsächlichen Fähigkeiten einer Person eine zentrale Rolle spielen, sondern die subjektiven Überzeugungen von diesen Fähigkeiten, selbst wenn diese nicht mit der Realität übereinstimmen (Brockmeyer & Edelstein, 1997, S. 17). Dies bedeutet, dass Wirksamkeitsüberzeugungen in erster Linie Wahrnehmungen widerspiegeln und nicht tatsächliche Kompetenzen. Brockmeyer und Edelstein (1997, S. 12) halten jedoch auch fest: „Stimmung, Motivation und Leistung werden indessen häufig stärker von Selbstwirksamkeitsüberzeugungen einer Person beeinflusst als vom tatsächlichen Fähigkeitsniveau.“ Trotzdem stellt sich die Frage, ob und wann Kompetenzwahrnehmungen revidiert werden. Dies hängt zum einen von den Kompetenzerfahrungen und zum anderen von der Ursachenzuschreibung des Individuums ab (vgl. z.B. Brockmeyer & Edelstein, 1997, S. 17; Schmitz & Schwarzer, 2000). Attribuiert eine Person demnach eine negative Selbstwirksamkeitserfahrung auf extrinsische, variable Faktoren, so geht dies nicht mit einer Veränderung der Selbstwirksamkeitserwartung einher. Attribuiert sie jedoch auf interne, stabile Faktoren, so wird sie ihre Selbstwirksamkeitserwartung revidieren.

Als Ursachen für die Entwicklung und auch die Veränderung von Selbstwirksamkeitserwartungen nennt Bandura (1997, S. 79) vier unterschiedlich effektive Quellen, die aufeinander aufbauen: So erhält eine Person durch körperliche Reaktionen (wie feuchte Hände, schneller Herzschlag etc.) Auskunft über die wahrgenommene Kompetenz. Verbale Mitteilungen anderer können ebenfalls das Gefühl von Selbstwirksamkeit aufbauen. Das Individuum kann durch Vorbilder lernen, dass bestimmte Situationen gemeistert werden können. Und schließlich ermöglicht eine direkte Erfahrung das Erleben von Selbstwirksamkeit. Je höher die Ebene, umso nachhaltiger wird Selbstwirksamkeit von den Individuen erlebt.

Selbstwirksamkeitsempfinden hängt damit auch von dem Verhalten anderer Personen ab. Fragen sie mich nach Rat? Unterstützen sie mich in meinen Ent-

scheidungen? Van Knippenberg, van Knippenberg, de Cremer und Hogg (2004) fanden beispielsweise heraus, dass im Arbeitskontext das Verhalten von Führungspersonen die wahrgenommene Selbstwirksamkeit der Mitarbeitenden beeinflusst. Brockmeyer und Edelstein (1997, S. 13) fordern dementsprechend für Organisationen Handlungsautonomie nach innen, Organisationsautonomie nach außen und eine kooperative, kommunikative und arbeitsteilige Struktur. Sie versprechen sich durch die Verwirklichung von Autonomie im Arbeitskontext, positive Selbstwirksamkeitserlebnisse bei den Mitarbeitenden fördern zu können.

Insgesamt zeigen bisherige Studien, dass die allgemeine Selbstwirksamkeit ein eher stabiles Konstrukt ist, wohingegen die spezifischen Selbstwirksamkeitserwartungen veränderlich erscheinen (Jonas & Brömer, 2002, zitiert nach Spieß & von Rosenstiel, 2010, S. 34; Schmitz, 1998, S. 148). Zwischen Faktoren wie Geschlecht, Berufserfahrung, Alter und Kompetenzerwartung liefern Studien bisher inkonsistente Ergebnisse, die besonders auch hinsichtlich der untersuchten Selbstwirksamkeitserwartung (allgemein versus bereichsspezifisch etc.) variieren (Schmitz, 1998, S. 147f.).

Als Effekte individueller Selbstwirksamkeit haben Studien u.a. die Reduzierung gesundheitlicher Probleme bei Schüler/-innen (Satow & Bäßler, 1998, S. 127), weniger Burnout bei Lehrpersonen sowie eine geringere Wahrnehmung von beruflicher Belastung und mehr Engagement im extracurricularen Unterricht (Schmitz & Schwarzer, 2000) aufgezeigt. Zudem fassen Brockmeyer und Edelstein (1997, S. 12) zusammen, dass eine hohe Selbstwirksamkeitserwartung in Organisationen das Organisationsklima sowie die Aktivität und Bereitschaft zur Verantwortungsübernahme günstig beeinflusst. Kolbe, Jerusalem und Mittag (1998, S. 124) stellen heraus, dass schulische Selbstwirksamkeitsüberzeugungen bei Schüler/-innen in einem positiven Zusammenhang zum Klassenklima stehen.

Hinsichtlich der kollektiven Selbstwirksamkeitserwartung stellt Bandura (1993, zitiert nach Schwarzer & Schmitz, 1999) die These auf, dass sich Lehrer/-innenkollegien, die sich als selbstwirksam wahrnehmen, unabhängig von Charakteristika der Schüler/-innen, in der Ausbildung der Schüler/-innen als erfolgreicher erweisen als Lehrpersonen an Schulen, deren Kollegium sich nicht als kollektiv wirksam erlebt. Schließlich heben auch Schwarzer und Schmitz (1999) die kollektive Selbstwirksamkeit als ein Element zur Prävention von Burnout bei Lehrerinnen hervor sowie als einen Faktor zur Unterstützung von Schulreformbewegungen und für die Umsetzung von Innovationsprozessen. Kollektive Selbstwirksamkeit sei demnach ein wichtiger Bestandteil der systemischen Schulberatung (ebd.).

## 2 Hypothesen

Im Folgenden werden die Hypothesen zu den drei eingangs gestellten Fragestellungen formuliert.

*Frage 1: Wie werden die Aspekte Schulklima, Arbeitszufriedenheit und individuelle sowie kollektive Selbstwirksamkeit von den Lehrpersonen über die Jahre wahrgenommen? Wie nehmen Schüler/-innen das Schulklima wahr?*

Diese Frage zielt auf die Beschreibung des wahrgenommenen Schulalltags und hat demnach explorativen Charakter. Es werden daher keine Hypothesen aufgestellt. Allerdings werden hinsichtlich der Einschätzung zur Arbeitszufriedenheit relativ hohe Zufriedenheitswerte erwartet (Amman, 2004; Grunder & Bieri, 1995; van Dick, 2006).

Hinsichtlich der Entwicklung über die Jahre wird angenommen, dass die Aspekte Schulklima und individuelle sowie kollektive Selbstwirksamkeit relativ stabile Konstrukte sind und sich daher über die Zeit wenig verändern (Buholzer, 2000; Jonas & Brömer, 2002, zitiert nach Spieß & von Rosenstiel, 2010, S. 34; Schmitz, 1998, S. 148; Ulich, 1998, S. 382, zitiert nach Varbelow, 2003).

In Bezug auf die Arbeitszufriedenheit, die laut Spieß und Winterstein (1999) stark von persönlichen und situativen Faktoren abhängt, werden über die Zeit stärkere Schwankungen erwartet. Trotzdem handelt es sich auch hier um langfristig gleichbleibende Bewertungen, weswegen im linearen Zeitverlauf keine starken Veränderungen angenommen werden.

*Frage 2: Zeigt sich ein Zusammenhang zwischen der Erfahrung der Lehrpersonen mit dem zentralen Prüfungssystem und den genannten Konstrukten?*

Hinsichtlich der Erfahrung mit dem Zentralabitur wird für das Konstrukt des Schulklimas kein Zusammenhang angenommen. Denn die Wahrnehmung des Schulklimas kann sich zwar verändern, da es jedoch kokonstruiert wird und das Schulklima Teil der Schulkultur ist (Bessoth & Weibel, 2003; Eder, 1996), wird angenommen, dass hier eher auf längere Sicht Wechsel eintreten. Kurzfristige, starke Belastungen (wie beispielsweise die Einführung zentraler Prüfungen) könnten zwar auch eine kurzzeitige Verschlechterung des Schulklimas zur Folge haben. Dies wird hier jedoch nicht angenommen, da durch das Schulklima in dieser Untersuchung besonders soziale Aspekte erfasst werden, die von den Prüfungen nicht oder weniger betroffen sind.

In Bezug auf die Arbeitszufriedenheit wird davon ausgegangen, dass die Reform sich negativ auswirken könnte (Bishop 1999; Spieß & Rosenstiel, 2010; van Dick, 2006). Denn das Zentralabitur als neues Instrument ist erst einmal mit zusätzlichen Belastungen (wie Umstellung der bisherigen Arbeitsweise etc.) ver-

bunden. Zudem engt es Handlungsspielräume der Lehrpersonen ein, was sich ebenfalls negativ auf die Arbeitszufriedenheit auswirken kann. In amerikanischen Studien wurde dieser negative Zusammenhang bereits festgestellt (Berryhill, Linney & Fromewick, 2009). Allerdings sind die Bedingungen, unter denen die Lehrpersonen in Deutschland arbeiten, nicht mit denen ihrer Kolleg/-innen in den USA vergleichbar. So sind in Deutschland die Ergebnisse der Schüler/-innen in den zentralen Prüfungen mit keinen negativen Konsequenzen für die Lehrpersonen verbunden.

Hinsichtlich der Selbstwirksamkeitsüberzeugungen der Lehrpersonen wird erwartet, dass durch die Einführung zentraler Prüfungen die individuelle wie auch die kollektive Selbstwirksamkeitserwartung geschwächt wird. Denn aufgrund dieses Instruments wird den Lehrpersonen Handlungsspielraum entzogen und dies kann das Erleben von Selbstwirksamkeit beeinträchtigen (Brockmeyer & Edelstein, 1997, S. 13).

*Frage 3: Stehen persönliche und biografische Faktoren im Zusammenhang mit den untersuchten Konstrukten?*

Als persönliche bzw. biografische Faktoren werden der Einfluss des Geschlechts, der Zugehörigkeit zu einem Bundesland und der Anzahl der Jahre im Schuldienst weitgehend explorativ untersucht. Denn Ergebnisse zu den untersuchten Konstrukten zeigen zumeist inkonsistente Befunde: So kommen empirische Studien hinsichtlich der Selbstwirksamkeitsüberzeugungen (Schmitz, 1998) und der Arbeitszufriedenheit zu uneinheitlichen Ergebnissen. Beispielsweise sind Frauen teilweise zufriedener mit der Arbeit als Männer (Merz, 1979b, 1980, zitiert nach van Dick, 2006, S. 126), teilweise finden sich aber auch keine Unterschiede (Ammann, 2004, S. 77ff.). Auch das Dienstalter steht teilweise in einem linearen und teilweise in einem U-förmigen Zusammenhang zur Arbeitszufriedenheit (Ammann, 2004, S. 77ff.).

In Bezug auf das Schulklima wird in der Literatur angenommen, dass es im Vergleich zu Faktoren wie der Lage und Größe der Schule oder des Schultyps weitgehend relativ invariant gegenüber persönlichen und biografischen Faktoren sei (Buholzer, 2000; Ulich, 1998, S. 382, zitiert nach Varbelow, 2003). Allerdings wird das Schulklima hier als personalistisches Konstrukt verstanden (Eder, 1996, S. 78), was bedeutet, dass es sich durch Erfahrungen und Wahrnehmung der Einzelnen und der Gruppe konstituiert. Diese Erfahrungen stehen wiederum in Bezug zu persönlichen und biografischen Faktoren wie z.B. dem Geschlecht. Es wird daher angenommen, dass persönliche und biografische Faktoren im Zusammenhang mit der Wahrnehmung des Schulklimas stehen. Hinsichtlich der weiteren Faktoren (Lage, Größe, etc.) wird ebenfalls ein Einfluss angenommen. Da es sich bei den untersuchten Schulen aber einzig um Gymnasien handelt und

zur Lage und Größe im Rahmen dieser Studien keine Daten erhoben wurden, werden diese Einflussfaktoren an dieser Stelle nicht berücksichtigt.

### 3 Design und Methoden

#### 3.1 Stichprobe

Es werden Daten aus den Jahren 2007, 2008 und 2009 von den in dieser Studie integrierten 19 Schulen in Bremen und 18 Schulen in Hessen berücksichtigt. Die Skala „individuelle Selbstwirksamkeit“ allerdings wurde im Jahr 2007 bei den Lehrpersonen nicht erfasst. Alle Lehrpersonen, also auch diejenigen, die nicht an zentralen Prüfungen beteiligt waren, wurden vor dem Abitur nach ihren Einschätzungen zu den Konstrukten befragt. Bei den Schüler/-innen wurden fehlende Werte imputiert, weswegen gepoolte Mittelwerte und Standardabweichungen angegeben werden. Bei den Lehrpersonen fand keine Imputation fehlender Werte statt (siehe hierzu den Beitrag von Maag Merki und Oerke in diesem Band: „3. Methodische Grundlagen der Studie“). Tabelle 1 zeigt an, wie viele Lehrpersonen und Schüler/-innen an der Befragung teilgenommen haben.

Tabelle 1: Stichprobe Lehrpersonen und Schüler/-innen (N)

	Lehrpersonen			Schüler/-innen		
	Hessen	Bremen	Summe	Hessen	Bremen	Summe
2007	1093	641	1734	973	751	1724
2008	949	577	1526	975	977	1952
2009	1005	424	1429	1017	1012	2029

Bei den Konstrukten, die in allen drei Jahren erhoben wurden (Schulklima, Arbeitsunzufriedenheit und kollektive Selbstwirksamkeit), liegen von 41% bis 42% der Lehrpersonen Antworten aus zwei Jahren vor (entspricht 1925 bis 1953 Personen). Von 13% bis 14% sind die Antworten aus drei Jahren erfasst (619 bis 638 Personen). In Bezug auf die Variable individuelle Selbstwirksamkeit, die nur 2008 und 2009 abgefragt wurde, liegen Daten von 30% der Lehrpersonen aus zwei Jahren vor (887 Personen).

#### 3.2 Messinstrumente

Im Folgenden werden als erstes die als abhängig eingeführten Variablen, als zweites die als unabhängig ausgewählten Variablen mit je einem Beispielitem und den Konsistenzwerten beschrieben.

- Schulklima Schüler/-innen (Skala; 9 Items; Cronbachs Alpha Hessen:  $\alpha_{2007} = .82$ ,  $\alpha_{2008} = .84$ ,  $\alpha_{2009} = .84$ ; Bremen:  $\alpha_{2007} = .86$ ,  $\alpha_{2008} = .84$ ,  $\alpha_{2009} = .84$ ). Beispielitem: „Die Stimmung an unserer Schule ist meistens... heiter, fröhlich – gedrückt, lustlos.“ Die verwendeten Attribute werden als Polaritäten vorgegeben und anhand einer fünfstufigen Skalierung erfasst (z.B. 5 = heiter, fröhlich; 1 = gedrückt, freudlos). Quelle: Eder (1998, S. 12); Leutwyler & Maag Merki (2004, S. 4).
- Schulklima Lehrpersonen (Skala; 9 Items; Cronbachs Alpha Hessen:  $\alpha_{2007} = .83$ ,  $\alpha_{2008} = .88$ ,  $\alpha_{2009} = .89$ ; Bremen:  $\alpha_{2007} = .85$ ,  $\alpha_{2008} = .83$ ,  $\alpha_{2009} = .88$ ). Die Items und das Antwortformat sind identisch zu denen der Schüler/-innen. Quelle: Eder (1998, S. 12) aus Halbheer, Kunz & Maag Merki (2005, S. 37).
- Individuelle Selbstwirksamkeit (Lehrpersonen) (Skala; 6 Items; nur 2008 und 2009 erhoben; Cronbachs Alpha Hessen:  $\alpha_{2008} = .73$ ,  $\alpha_{2009} = .78$ ; Bremen:  $\alpha_{2008} = .76$ ,  $\alpha_{2009} = .79$ ). Beispielitem: „Ich bin mir sicher, dass ich auch mit den problematischsten Schülern und Schülerinnen in guten Kontakt kommen kann, wenn ich mich darum bemühe.“ Antwortformat: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu. Quelle: Schwarzer & Jerusalem (1999)<sup>1</sup>.
- Kollektive Selbstwirksamkeit (Lehrpersonen) (Skala; 5 Items; Cronbachs Alpha Hessen:  $\alpha_{2007} = .67$ ,  $\alpha_{2008} = .77$ ,  $\alpha_{2009} = .75$ ; Bremen:  $\alpha_{2007} = .69$ ,  $\alpha_{2008} = .73$ ,  $\alpha_{2009} = .72$ ). Beispielitem: „Auch mit außergewöhnlichen Vorfällen können wir zurechtkommen, da wir uns im Kollegium gegenseitig Rückhalt bieten.“ Antwortformat: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu. Quelle: Schwarzer & Jerusalem (1999).
- Arbeitsunzufriedenheit (Lehrpersonen) (Skala; 6 Items; Cronbachs Alpha Hessen:  $\alpha_{2007} = .84$ ,  $\alpha_{2008} = .81$ ,  $\alpha_{2009} = .82$ ; Bremen:  $\alpha_{2007} = .83$ ,  $\alpha_{2008} = .79$ ,  $\alpha_{2009} = .82$ ). Beispielitem: „Ich habe mir schon ernsthaft überlegt, aus dem Beruf auszusteigen.“ Antwortformat: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu. Quelle: Schwarzer & Jerusalem (1999), Halbheer, Kunz & Maag Merki (2005, S. 58).
- Bundesland: 1 = Bremen; 0 = Hessen.
- Erfahrung mit dem Zentralabitur (Lehrpersonen): 0 = keine Erfahrung, 1 = ein Jahr Erfahrung, 2 = zwei Jahre Erfahrung, 3 = drei Jahre Erfahrung.
- Geschlecht (Lehrpersonen): 1 = weiblich, 0 = männlich.

<sup>1</sup> Skalen zur Erfassung von Lehrer/-innen- und Schüler/-innenmerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen.



- Jahr (Lehrpersonen): Zum einen wurde die Variable Jahr als intervallskalierte Variable eingeführt: 2007 = 0; 2008 = 1; 2009 = 2. Durch sie werden lineare Effekte von 2007 zu 2009 längsschnittlich untersucht. Zum anderen wurden zur Analyse der Effekte von 2007 zu 2008 und 2008 zu 2009 zwei Dummies gebildet: Jahr07: 1 = 2007 und Jahr09: 1 = 2009. Die Referenz ist jeweils das Jahr 2008.
- Jahr (Schüler/-innen): Zur Analyse der Effekte von 2007 zu 2008 und 2007 zu 2009 wurden zwei Dummies gebildet: Jahr08: 1 = 2008 und Jahr09: 1 = 2009. Die Referenz ist jeweils das Jahr 2007.
- Schuldienst (Lehrpersonen): Da die Variable „Anzahl der Jahre im Schuldienst“ ordinal skaliert ist, wurde sie zu zwei Dummy-Variablen zusammengefasst, die jeweils auf die Kategorie referieren, die der durchschnittlichen Anzahl der Dienstjahre der Lehrpersonen an einer Schule entspricht (6-20 Jahre). Die erste Dummy-Variable „Schuldienst kurz“ (Schuldienst\_k) umfasst damit alle Lehrpersonen, die Referendar/-innen sind oder maximal fünf Jahre an einer Schule arbeiten (= 1). Die zweite Dummy-Variable „Schuldienst lang“ (Schuldienst\_l) berücksichtigt alle Lehrpersonen, die 21 Jahre und mehr im Schuldienst stehen (= 1).

### 3.3 Auswertungsverfahren

Als abhängige Variable wurden bei den Lehrpersonen das Schulklima, die Arbeitsunzufriedenheit, die individuelle sowie die kollektive Selbstwirksamkeit untersucht. Hinsichtlich der Wahrnehmung des Schulklimas wurde zudem die Sicht der Schüler/-innen dargestellt.

Die Daten der Schüler/-innen und Lehrpersonen wurden pro Jahr deskriptiv ausgewertet. Um Unterschiede zwischen den Jahren innerhalb eines Bundeslandes bei den Schüler/-innen zu untersuchen, wurde eine Regressionsanalyse durchgeführt. Die Angaben der Lehrpersonen wurden zudem mehr Ebenenanalytisch mit dem Software Programm HLM (Hierarchical Linear and Nonlinear Modeling) von Raudenbush, Bryk und Congdon (2004) in zwei Modellen untersucht, die beide je drei Ebenen umfassen: Die erste Ebene beinhaltet die intraindividuellen Veränderungen über die Jahre und die zweite Ebene die Unterschiede zwischen den Individuen. Die dritte Ebene umfasst die Differenzen zwischen den Schulen. Es wurden für die Konstrukte Schulklima, Arbeitsunzufriedenheit und kollektive Selbstwirksamkeit zwei Modelle berechnet, um so Effekte über die Jahre von 2007-2009 und zwischen den Jahren 2007-2008 und 2008-2009 zu analysieren. Dies hat den Vorteil, dass neben linearen Effekten auch quadratische bzw. U-förmige Effekte untersucht werden können. Da in Bezug auf die indivi-

duelle Selbstwirksamkeit nur Daten aus den Jahren 2008 und 2009 vorliegen, wurde dieses Konstrukt nur in der längsschnittlichen Analyse berücksichtigt.

Alle vier abhängigen Variablen sind laut Shapiro-Wilk und dem Kolmogorov-Smirnov-Test nicht normal verteilt. Berry und Feldman (1985, S. 11) sowie Bohrnstedt und Carter (1971) argumentieren allerdings, dass dies in Regressionen nur bei kleinen Fallzahlen zu relevanten Verzerrungen der Signifikanzwerte führt. Zhang (2005, S. 64) weist darauf hin, dass Verletzungen der Normalverteilungsannahme auch bei hierarchischen Mehrebenenanalysen keine relevanten Verzerrungen der Ergebnisse zur Folge habe. Auf die Transformation der abhängigen Variablen wurde daher verzichtet und es wurden parametrische Verfahren angewendet. Es sind in den Tabellen die unstandardisierten Effekte und fixen Effekte mit robusten Standardfehlern aufgeführt.

Im Folgenden sind die eingesetzten Regressionsgleichungen angegeben, wobei die Variablen kursiv gedruckt sind, die zentriert (grand mean) eingegeben wurden.

Schulklima/ Arbeitsunzufriedenheit/ Individuelle Selbstwirksamkeit /Kollektive Selbstwirksamkeit<sub>2007-2009</sub> =  $\gamma_{000} + \gamma_{001} * \text{Bremen} + \gamma_{010} * \text{Geschlecht} + \gamma_{100} * \text{ErfZentral} + \gamma_{200} * \text{Jahr} + \gamma_{300} * \text{Schuldienst\_k} + \gamma_{400} * \text{Schuldienst\_l} + r_0 + u_{00} + e$

Schulklima/ Arbeitsunzufriedenheit/Kollektive Selbstwirksamkeit<sub>2007-2008, 2008-2009</sub> =  $\gamma_{000} + \gamma_{001} * \text{Bremen} + \gamma_{010} * \text{Geschlecht} + \gamma_{100} * \text{ErfZentral} + \gamma_{200} * \text{Jahr07} + \gamma_{300} * \text{Jahr09} + \gamma_{400} * \text{Schuldienst\_k} + \gamma_{500} * \text{Schuldienst\_l} + r_0 + u_{00} + e$

#### 4 Ergebnisse

*Frage 1: Wie werden die Aspekte Schulklima, individuelle und kollektive Selbstwirksamkeit sowie Arbeitszufriedenheit von den Lehrpersonen über die Jahre wahrgenommen? Wie nehmen Schüler/-innen das Schulklima wahr?*

Tabelle 2 zeigt, dass die hessischen wie die Bremer Lehrpersonen in 2007, 2008 und 2009 durchschnittlich das Schulklima als eher bis sehr positiv bewerten. So geben zwischen 83% und 88% der Lehrpersonen in den drei untersuchten Jahren an, dass sie sich an der Schule persönlich behandelt fühlen und/oder das Klima als fröhlich und heiter wahrnehmen (Werte > 3.5). In allen drei Jahren sagt maximal 1% der Lehrpersonen, dass sie das Schulklima als eher negativ erleben (Werte < 2.5).

Ihren Berufsalltag beurteilen die Lehrpersonen ebenfalls als eher positiv. So stimmen zwischen 82% und 88% der Lehrpersonen in Hessen bzw. Bremen in 2007, 2008 und 2009 gar nicht bzw. eher nicht zu, dass sie mit ihrer Arbeit unzufrieden seien (Werte > 2.5).

Hinsichtlich ihrer individuellen Selbstwirksamkeit fühlen sich zwischen 69% und 75% der Lehrpersonen in 2008 und 2009 als eher selbstwirksam ( $2.5 < \text{Werte} < 3.5$ ). In Bremen geben 8% und in Hessen zwischen 10% und 15% sogar an, dass sie (beispielsweise) der Aussage voll zustimmen, dass sie sich sicher sind, selbst den problematischsten Schüler/-innen den prüfungsrelevanten Stoff vermitteln zu können ( $\text{Werte} > 3.5$ ). Zwischen 9% und 22% der Lehrer/-innen sind sich eher nicht sicher, ob sie dies bewerkstelligen können ( $\text{Werte} < 2.5$ ).

Tabelle 2: Deskriptive Auswertungen: Lehrpersonen

		Schulklima	Arbeitsunzufriedenheit	Individuelle Selbstwirksamkeit <sup>a)</sup>	Kollektive Selbstwirksamkeit
<b>Hessen</b>					
2007	M	3.99	1.84	-	2.57
	SD	0.52	0.54	-	0.49
	N	1059	1052	-	1047
2008	M	3.95	1.86	2.86	2.58
	SD	0.53	0.50	0.43	0.52
	N	919	913	904	907
2009	M	3.97	1.79	2.96	2.64
	SD	0.50	0.50	0.45	0.49
	N	994	993	983	987
<b>Bremen</b>					
2007	M	4.07	1.91	-	2.53
	SD	0.52	0.56	-	0.49
	N	614	611	-	610
2008	M	4.05	1.91	2.81	2.49
	SD	0.50	0.51	0.46	0.49
	N	572	561	557	561
2009	M	4.04	1.92	2.76	2.47
	SD	0.51	0.55	0.48	0.49
	N	424	422	420	420

M = Mittelwert; SD = Standardabweichung; N = Anzahl Fälle; Antwortformat Schulklima: Die verwendeten Attribute werden als Polaritäten vorgegeben und anhand einer fünfstufigen Skalierung erfasst (z.B. 5 = heiter, fröhlich; 1 = gedrückt, freudlos); Arbeitsunzufriedenheit, individuelle und kollektive Selbstwirksamkeit: 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu a) Die Skala Individuelle Selbstwirksamkeit wurde nur 2008 und 2009 erhoben.

Hinsichtlich der wahrgenommenen kollektiven Selbstwirksamkeit liegt der Anteil derer, die das Kollegium als eher selbstwirksam erleben in den drei Jahren in Hessen zwischen 58% und 64% und in Bremen zwischen 47% und 50% ( $\text{Werte} > 2.5$ ). Zwischen 34% (Hessen 2009) und 51% der Lehrpersonen (Bremen

2009) sehen das Kollegium als eher nicht selbstwirksam ( $1.5 < \text{Werte} < 2.5$ ). Zwischen 1% bis 3% der Lehrer/-innen hat gar nicht das Gefühl, dass sein bzw. ihr Kollegium selbstwirksam sei und Herausforderungen auch bei besonderen Schwierigkeiten meistern könne (Werte  $< 1.5$ ). In Bezug auf die kollektive wie die individuelle Selbstwirksamkeit schätzen sich in allen drei Jahren die Hessischen Lehrpersonen selbstwirksamer ein als die Bremer Lehrpersonen (vgl. Tabelle 2).

Auch die Schüler/-innen erleben das Schulklima an ihren Schulen als eher positiv (vgl. Tabelle 3). Diese Einschätzung teilen die Hessischen mit den Bremer Schüler/-innen, es liegen in keinem Jahr signifikante Unterschiede zwischen den Bundesländern vor. Zudem zeigen sich auch keine signifikanten Unterschiede zwischen den Jahren innerhalb der einzelnen Bundesländer (Regressionsanalyse: Jahresvergleich).

*Tabelle 3: Deskriptive Auswertungen: Schüler/-innen*

	2007	Schulklima 2008	2009
<b>Hessen</b>			
M	3.53	3.50	3.51
SD	0.54	0.56	0.57
SE	0.02	0.02	0.02
N	1038	1067	1095
<b>Bremen</b>			
M	3.53	3.49	3.51
SD	0.62	0.58	0.57
SE	0.02	0.02	0.02
N	959	1052	1120

M = kombinierter Mittelwert; SD = kombinierte Standardabweichung (naïves Pooling); SE = Standardfehler Mittelwertschätzung; N = Anzahl Fälle; Antwortformat Schulklima: Die verwendeten Attribute werden als Polaritäten vorgegeben und anhand einer fünfstufigen Skalierung erfasst (z.B. 5 = heiter, fröhlich; 1 = gedrückt, freudlos)

Um die Entwicklung der Wahrnehmungen der Lehrpersonen über die Jahre unter Kontrolle verschiedener Prädiktoren zu untersuchen, wurden mehrbenenanalytische Auswertungen einmal in Bezug auf lineare Effekte (vgl. Tabelle 4) und einmal hinsichtlich der Effekte von 2007 zu 2008 und 2008 zu 2009 durchgeführt (vgl. Tabelle 5).

Hinsichtlich des Schulklimas und der Arbeitsunzufriedenheit zeigen sich in Bezug auf lineare Entwicklungen keine signifikanten Effekte (vgl. Tabelle 4). Betrachtet man die Entwicklung von 2007 zu 2008 und von 2008 zu 2009 (vgl. Tabelle 5), so zeigt sich, dass sich das Schulklima nicht verändert. Bei der Arbeitsunzufriedenheit lässt sich von 2007 zu 2008 ein leichter, nicht signifikan-

ter Anstieg verzeichnen, von 2008 zu 2009 ist er wieder rückläufig ( $p < .01$ ). Die Veränderung ist aber mit einem standardisiertem Koeffizienten von -0.03 gering (vgl. Tabelle 6). Die Einschätzungen in 2007 und 2009 unterscheiden sich so-dann auch nicht (vgl. Tabelle 4).

Tabelle 4: Mehrebenenanalysen: 2007-2009

Fixe Effekte	SK	AUZ	IS <sup>a)</sup>	KS
<b>Ebene 1</b>				
Konstante, $B_0$	3.92 (0.04)***	1.89 (0.03)***	2.86 (0.03)***	2.54 (0.03)***
ErfZentral	n.s.	0.02 (0.01) <sup>+</sup>	-0.04 (0.01)***	-0.04 (0.01)***
Jahr	n.s.	n.s.	0.07 (0.02)***	0.03 (0.01)*
Schuldienst_k	0.06 (0.02)**	-0.09 (0.02)***	0.09 (0.02)**	0.09 (0.02)***
Schuldienst_l	n.s.	n.s.	n.s.	0.05 (0.02)*
<b>Ebene 2</b>				
Geschlecht (weiblich = 1)	0.06 (0.02)**	-0.06 (0.02)**	0.04 (0.02) <sup>+</sup>	n.s.
<b>Ebene 3</b>				
Bundesland (Bremen = 1)	0.09 (0.05) <sup>+</sup>	0.08 (0.03)**	-0.11 (0.03)***	-0.08 (0.03)*
<b>Zufällige Effekte</b>				
Varianz- komponenten				
E3: $u_{00}(\tau_{\beta 0})$	0.02***	0.00*	0.00*	0.01***
E2: $\tau_0(\tau_{\pi 0})$	0.15***	0.20***	0.13***	0.12***
E1: $e(\sigma^2)$	0.09	0.08	0.06	0.11
Devianz	5200	5186	2729	5263
N der geschätz- ten Parameter	10	10	10	10
N Ebene 1	4239	4234	2625	4204
N Ebene 2	2458	2461	1845	2446
N Ebene 3	37	37	37	37

Schätzer: unstandardisierte Koeffizienten (Standardfehler); E = Ebene; <sup>+</sup>  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; n.s. = nicht signifikant; Antwortformat Schulklima (SK): Die verwendeten Attribute werden als Polaritäten vorgegeben und anhand einer fünfstufigen Skalierung erfasst (z.B. 5 = heiter, fröhlich; 1 = gedrückt, freudlos); Arbeitsunzufriedenheit (AUZ), individuelle/kollektive Selbstwirksamkeit (IS/KS): 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu; a) Die Skala Individuelle Selbstwirksamkeit wurde nur 2008 und 2009 erhoben.

In Bezug auf das Schulklima können 37% der Varianz durch intraindividuelle Faktoren, 55% durch Differenzen zwischen den Lehrpersonen und 8% durch die Schulebene erklärt werden. 70% der Unterschiede in der Arbeitsunzufriedenheit lassen sich durch Differenzen zwischen den Individuen erklären, 29% der Varianz gehen auf intrapersonelle Faktoren zurück und lediglich 1% der Unterschiede basieren auf Differenzen zwischen den Schulen.

Tabelle 5: Mehrebenenanalysen: 2007 – 2008, 2008 - 2009

Fixe Effekte	SK	AUZ	KS
<i>Ebene 1</i>			
Konstante, B <sub>0</sub>	3.91 (0.05)***	1.91 (0.03)***	2.53 (0.03)***
ErfZentral	n.s.	0.02 (0.01) <sup>+</sup>	-0.04 (0.01)***
Jahr07	n.s.	n.s.	n.s.
Jahr09	n.s.	-0.04 (0.01)**	0.05 (0.02)*
Schuldienst_k	0.06 (0.02)**	-0.09 (0.02)***	0.09 (0.02)***
Schuldienst_l	n.s.	n.s.	0.05 (0.02)*
<i>Ebene 2</i>			
Geschlecht (weiblich = 1)	0.06 (0.02)**	-0.06 (0.02)**	n.s.
<i>Ebene 3</i>			
Bundesland (Bremen = 1)	0.09 (0.05) <sup>+</sup>	0.08 (0.03)**	-0.08 (0.03)*
<i>Zufällige Effekte</i>			
Varianz- komponenten			
E3: $u_{00}(\tau_{p0})$	0.02***	0.00*	0.01***
E2: $r_0(\tau_{\pi 0})$	0.15***	0.20***	0.12***
E1: $e(\sigma^2)$	0.09	0.08	0.11
Devianz	5199	5178	5261
N der geschätz- ten Parameter	11	11	11
N Ebene 1	4239	4234	4204
N Ebene 2	2458	2461	2446
N Ebene 3	37	37	37

Schätzer: unstandardisierte Koeffizienten (Standardfehler); <sup>+</sup>  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; n.s. = nicht signifikant; Antwortformat Schulklima (SK): Die verwendeten Attribute werden als Polaritäten vorgegeben und anhand einer fünfstufigen Skalierung erfasst (z.B. 5 = heiter, fröhlich; 1 = gedrückt, freudlos); Arbeitsunzufriedenheit (AUZ), kollektive Selbstwirksamkeit (KS): 1 = trifft gar nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft genau zu

Positive lineare Entwicklungen über die Zeit sind in Bezug auf die individuelle (standardisierter Koeffizient: 0.13,  $p < .001$ ) und die kollektive Selbstwirksamkeit (0.05,  $p < .001$ ) zu beobachten (vgl. Tabelle 4 und Tabelle 6). Da für die individuelle Selbstwirksamkeit nur Daten aus 2008 und 2009 vorliegen, gibt es keine weiteren Auswertungen zu den Jahren. In Bezug auf die kollektive Selbstwirksamkeit bestätigen die Analysen zu den Effekten von 2007 zu 2008 und 2008 zu 2009, dass ein linearer Trend vorliegt: Die Wahrnehmung der kollektiven Selbstwirksamkeit nahm von 2007 zu 2008 leicht zu (wenn auch nicht signifikant) und ebenfalls von 2008 zu 2009 ( $p < .05$ ) (vgl. Tabelle 5). Die Auswertungen zeigen jedoch auch, dass der Effekt weitgehend auf die Veränderung von 2008 zu 2009 zurückgeht und relativ gering ist (2008 zu 2009: stand. Koef-

fizient = 0.04; vgl. Tabelle 6). Auch bei den Selbstwirksamkeitsüberzeugungen liegt ein Großteil der zu erklärenden Varianz auf der Ebene der Individuen: 64% bei der Wahrnehmung der individuellen und 50% bei der Wahrnehmung der kollektiven Selbstwirksamkeit. Intrapersonelle Veränderungen erklären 33% der Varianz der individuellen und 46% der Varianz der kollektiven Selbstwirksamkeit. Nur 3% beider Konstrukte variieren aufgrund von Faktoren auf der Schulebene.

Tabelle 6: Standardisierte Koeffizienten

	SK	AUZ	IS <sup>a)</sup>	KS
<i>Ebene 1</i>				
ErfZentral	n.s.	0.03	-0.08	-0.07
Jahr	n.s.	n.s.	0.13	0.05
Jahr07 <sup>b)</sup>	n.s.	n.s.	-	n.s.
Jahr09 <sup>b)</sup>	n.s.	-0.03	-	0.04
Schuldienst_k	0.05	-0.08	0.08	0.08
Schuldienst_l	n.s.	n.s.	n.s.	0.05
<i>Ebene 2</i>				
Geschlecht (weiblich = 1)	0.06	-0.05	0.05	n.s.
<i>Ebene 3</i>				
Bundesland (Bremen = 1)	0.09	0.08	-0.12	-0.08

Standardisierung nach Hox (2002, S. 21); n.s. = nicht signifikant; Schulklima (SK); Arbeitsunzufriedenheit (AUZ); individuelle/kollektive Selbstwirksamkeit (IS/KS); a) Die Skala Individuelle Selbstwirksamkeit wurde nur 2008 und 2009 erhoben; b) Da die unstandardisierten Koeffizienten sich nahezu entsprechen (vgl. Tabelle 4 und 5), wurde auf eine zweifache Standardisierung der Werte verzichtet. Die standardisierten Werte beruhen – abgesehen von den Dummies Jahr07 und Jahr09 – auf der Mehrebenenanalyse 2007 und 2009 (vgl. Tabelle 4). Die standardisierten Werte der Dummies Jahr07 und Jahr09 stammen aus der Mehrebenenanalyse 2007-2008 und 2008-2009 (vgl. Tabelle 5).

*Frage 2: Zeigt sich ein Zusammenhang zwischen der Erfahrung der Lehrpersonen mit dem zentralen Prüfungssystem und den genannten Konstrukten?*

In Tabelle 4 sieht man, dass sich in Bezug auf das Schulklima kein Zusammenhang zu der Erfahrung mit den zentralen Prüfungen zeigt. Hinsichtlich der Arbeitsunzufriedenheit ist ein leichter Anstieg der Unzufriedenheit (signifikant auf dem 10%-Niveau) zu beobachten (standardisierter Koeffizient: 0.03; vgl. Tabelle 6). In Bezug auf die Wahrnehmung der individuellen und kollektiven Selbstwirksamkeit ist jeweils ein negativer Zusammenhang sichtbar: Die Selbstwirksamkeitserwartung sinkt mit zunehmender Erfahrung mit dem zentralen Prüfungssystem ( $p < .001$ ). Die standardisierten Koeffizienten betragen -0.07 (kollektive) und -0.08 (individuelle Selbstwirksamkeit) (vgl. Tabelle 6).

*Frage 3: Stehen persönliche und biografische Faktoren im Zusammenhang mit den untersuchten Konstrukten?*

Hinsichtlich der Einflussfaktoren zeigen sich mit den untersuchten Konstrukten relativ einheitliche Ergebnisse in Bezug auf die kurze Zeit im Schuldienst (< 6 Jahre) und in Bezug auf das Geschlecht: Lehrpersonen, die weniger als sechs Jahre im Schuldienst und/oder weiblich sind, nehmen das Schulklima als positiver, die individuelle Selbstwirksamkeit als höher wahr und sind mit ihrer Arbeit zufriedener als Kolleg/-innen, die mehr als 6 Jahre im Schuldienst und/oder männlich sind.

Bei den Lehrpersonen, die länger als 21 Jahre im Schuldienst sind, zeigt sich lediglich in Bezug auf die kollektive Selbstwirksamkeit ein positiver signifikanter Effekt: Lehrpersonen, die lange im Schuldienst sind, nehmen – vergleichbar mit den Lehrpersonen, die weniger als 6 Jahre im Schuldienst stehen – eine höhere Selbstwirksamkeit des Kollegiums wahr als Lehrpersonen, die zwischen 6 und 20 Jahre an Schulen arbeiten. Zudem gibt es keinen signifikanten Unterschied zwischen Männern und Frauen in der Wahrnehmung der kollektiven Selbstwirksamkeit.

In Bezug auf den Einfluss des Bundeslandes sind keine einheitlichen Ergebnisse zu beobachten, allerdings wird der Zusammenhang zwischen der Zugehörigkeit zu einem Bundesland und allen untersuchten, abhängigen Variablen signifikant. Die Bremer bewerten das Schulklima als positiver als die hessischen Kolleg/-innen, dagegen sind die Lehrpersonen aus Hessen zufriedener mit ihrer Arbeit und nehmen sich selbst sowie das Kollegium als selbstwirksamer wahr als die Bremer Lehrpersonen.

## **5 Diskussion**

### *5.1 Schulklima*

Das Schulklima wird von 83% bis 88% der Lehrpersonen in Hessen und Bremen als eher bis sehr positiv bewertet. Bezüglich der Entwicklung über die drei Jahre zeigt sich, wie erwartet, kein linearer, signifikanter Zusammenhang. Von 2007 zu 2009 sowie von 2007 zu 2008 und 2008 zu 2009 blieb die Wahrnehmung des Schulklimas konstant. Unterschiede zwischen den Lehrpersonen erklären einen Großteil der Varianz in der Wahrnehmung des Schulklimas (55%). Allerdings kann auch ein beachtlicher Anteil (8%) durch die Zugehörigkeit zu einer Schule erklärt werden. Dies entspricht den Erwartungen und den bisherigen empirischen Befunden (Buholzer, 2000).

Um das Schulklima, welches subjektiv und kollektiv konstruiert wird, besser beschreiben zu können, wurde hier auch die Einschätzung der Schüler/-innen



berücksichtigt („multi-source Ansatz“, Fend, 1998, S. 48). Es zeigt sich, dass die Schüler/-innen das Schulklima über alle drei Jahre und in Hessen wie in Bremen gleichermaßen als (eher) positiv einschätzen. Die Lehrpersonen nehmen das Schulklima positiver wahr als die Schüler/-innen. Dies könnte dadurch erklärt werden, dass die Wahrnehmung des Klimas zwischen den Hierarchieebenen stark variiert, wie Ergebnisse aus der Wirtschaft zeigen (Payne & Pugh, 1983, S. 1166, referiert nach Bessoth & Weibel, 2003, S. 42): Führungskräfte beurteilen das Klima zumeist positiver als das „operative Personal“ (ebd.). Es könnte sinnvoll sein, in zukünftigen Studien zu untersuchen, warum die Schüler/-innen das Klima negativer bewerten als die Lehrpersonen. So würde das Ergebnis im Sinne Fends (2006) als Indikator für Problemkonstellationen der Schule verwendet werden.

Die Erfahrung mit dem Zentralabitur steht in keinem Zusammenhang mit der Beurteilung des Schulklimas. Dies entspricht allgemein den Erwartungen, da die zentralen Prüfungen nicht direkt das soziale Leben der Lehrer/-innen im Schulalltag betreffen.

Das Schulklima, welches hier als personalistisches Konstrukt verstanden wird, steht im Zusammenhang mit persönlichen Faktoren: So nehmen Frauen (standardisierter Koeffizient: 0.06), Lehrpersonen, die weniger als 6 Jahre an einer Schule unterrichten (standardisierter Koeffizient: 0.05) und Bremer Lehrpersonen (standardisierter Koeffizient: 0.09) das Schulklima positiver wahr als Männer, Lehrpersonen, die seit mehr als 6 Jahren an einer Schule arbeiten und hessische Lehrpersonen. Damit bestätigt sich die Hypothese, dass die Wahrnehmung des Schulklimas mit persönlichen bzw. biografischen Faktoren im Zusammenhang steht.

## 5.2 *Arbeitsunzufriedenheit*

In den drei Jahren haben zwischen 84% und 88% der Lehrpersonen angegeben, sie seien mit ihrer Arbeit „nicht“ bzw. „eher nicht“ unzufrieden. Dies wurde aufgrund der generell hohen Zustimmungsraten zur Arbeitszufriedenheit unter Lehrpersonen auch angenommen (Ammann, 2004; Grunder & Bieri, 1995; van Dick, 2006). Fraglich bleibt damit auch in dieser Studie, ob das Instrument zur Messung der Arbeitszufriedenheit angemessen ist. Denn aufgrund der speziellen Situation des Lehrberufs in Deutschland (verbeamtet und weitgehend isolierte, eigenverantwortliche Tätigkeit) sollten vielleicht weniger allgemeine Aussagen zur Arbeitszufriedenheit erfasst werden, sondern eher Einstellungen zu spezifischen „arbeitsrelevanten Gegebenheiten“, wie Spieß und Winterstein (1999) vorschlagen. Beispielsweise bezieht sich eine Aussage der hier präsentierten Skala auf die „ernsthafte Überlegung“, „den Beruf zu wechseln“. Dies ist jedoch in

Deutschland kaum eine Option für Lehrpersonen, da es sehr schwer ist, in anderen Bereichen Fuß zu fassen. Zudem wäre eine solche Entscheidung u.U. mit dem Verlust von Privilegien (z.B. gesicherter Arbeitsplatz, Verbeamtung, sichere Rente) verbunden. Die Ergebnisse zur Arbeitsunzufriedenheit geben damit nur einen ersten, allgemeinen Anhaltspunkt.

Unterschiede in der Arbeitsunzufriedenheit gehen zu 70% auf Differenzen zwischen den Individuen zurück, Schulunterschiede spielen keine Rolle (1% erklärte Varianz). Die restlichen 29% werden durch intraindividuelle Varianzen erklärt.

Über die drei untersuchten Jahre bleiben die Ergebnisse der Befragung zur Arbeitsunzufriedenheit stabil. Es zeigt sich aber bei den Analysen der Unterschiede von 2008 zu 2009 ein signifikanter Rückgang der Unzufriedenheit, der allerdings gering ist (standardisierter Koeffizient: -0.03) und sich im linearen Verlauf nicht abbildet. Dies entspricht der Erwartung, dass die Arbeitsunzufriedenheit langfristige Bewertungen umfasst und gleichzeitig situativen Schwankungen unterworfen ist.

Hinsichtlich der Erfahrung mit dem Zentralabitur bestätigt sich die Erwartung, dass ein negativer Zusammenhang zwischen der Erfahrung mit dem Zentralabitur und der Zufriedenheit im Beruf zu beobachten ist: Je mehr Erfahrung eine Lehrperson mit dem zentralen Prüfungssystem hat, desto unzufriedener ist sie. Dies, so wird vermutet, liegt daran, dass das Zentralabitur die Autonomie der Lehrpersonen einschränkt (Spieß & Rosenstiel, 2010; van Dick, 2006). Der Effekt ist allerdings klein (standardisierter Koeffizient: 0.03) und nur auf dem 10%-Niveau signifikant.

Hinsichtlich des Einflusses der persönlichen Faktoren auf die Arbeitszufriedenheit zeigt sich, dass Frauen zufriedener sind als Männer (standardisierter Koeffizient: -0.03) und dass Lehrpersonen, die weniger als 6 Jahre unterrichten, in ihrem Beruf zufriedener sind als ihre Kolleg/-innen mit einer Dienstzeit von über 6 Jahren (standardisierter Koeffizient: -0.08). Schließlich sind die hessischen Lehrenden etwas zufriedener als ihre Bremer Kolleg/-innen (standardisierter Koeffizient: 0.08).

### 5.3 Individuelle Selbstwirksamkeit

Knapp drei Viertel der Lehrpersonen empfinden sich als selbstwirksam. Allerdings geben auch zwischen 9% und 22% der Lehrer/-innen an, dass sie sich eher nicht sicher sind, ob sie unter schwierigen Bedingungen Herausforderungen bewältigen können. Zwischen 2008 und 2009 ist ein leichter, signifikanter Anstieg der Selbstwirksamkeitserwartung zu verzeichnen. Aspekte der Schule sind für die unterschiedliche Wahrnehmung nicht so relevant (3% der erklärten Varianz).

Stattdessen sind Unterschiede zwischen den Individuen (64%) und intrapersonelle Unterschiede ausschlaggebend.

Die Hypothese, dass die Erfahrung mit den zentralen Prüfungen negativ mit der wahrgenommenen individuellen Selbstwirksamkeit in Beziehung steht, kann bestätigt werden ( $p < .001$ ). Dieser Effekt ist mit einem standardisierten Koeffizienten von  $-0.08$  relativ gering. Er kann jedoch als Hinweis darauf gewertet werden, dass der Kontrollverlust durch die zentralen Prüfungen sich auf die Wahrnehmung der individuellen Selbstwirksamkeit auswirken könnte.

Hinsichtlich persönlicher Faktoren zeigt die vorliegende Studie, dass die individuelle Selbstwirksamkeitserwartung von den Frauen ( $p < .10$ ; standardisierter Koeffizient:  $0.05$ ) und von den hessischen Lehrpersonen ( $p = .001$ ; standardisierter Koeffizient:  $-0.12$ ) höher eingeschätzt wird als von den Männern und den Lehrpersonen in Bremen. Lehrpersonen, die weniger als sechs Jahre an einer Schule beschäftigt sind, nehmen sich selbst ebenfalls als selbstwirksamer wahr ( $p < .01$ ; standardisierter Koeffizient:  $0.08$ ). Hier müssten weitere Analysen zeigen, welche Ursachen für die unterschiedliche Einschätzung der Selbstwirksamkeit verantwortlich sind (z.B. höhere Kompetenzen, bessere Ausbildung).

#### 5.4 Kollektive Selbstwirksamkeit

Im Vergleich zur individuellen Selbstwirksamkeitserwartung liegt der Anteil der Lehrpersonen, die das Kollegium als selbstwirksam beschreiben, niedriger: In Hessen zwischen 58% und 64% und in Bremen zwischen 47% und 50% der Lehrpersonen. Knapp die Hälfte der Lehrpersonen sieht das Kollegium als eher nicht selbstwirksam. Über die drei Jahre zeigt sich ein leichter, signifikanter linearer Anstieg in der Wahrnehmung der Selbstwirksamkeit des Kollegiums. Dieser ist besonders auf den Zuwachs von 2008 zu 2009 zurückzuführen.

50% der Varianz in der Wahrnehmung kollektiver Selbstwirksamkeit kann durch Unterschiede zwischen den Individuen erklärt werden, doch auch die intrapersonellen Faktoren sind fast ebenso relevant: Sie erklären 46% der Varianz. Die Schule dagegen spielt als Faktor eine untergeordnete Rolle (3% erklärte Varianz).

Die Erfahrung mit dem zentralen Prüfungssystem steht, wie erwartet, in einem negativen Zusammenhang mit der wahrgenommenen kollektiven Selbstwirksamkeit ( $p = .001$ ). Dieser Effekt ist allerdings relativ klein (standardisierter Koeffizient:  $-0.07$ ). Damit ist in Ansätzen die Hypothese bestätigt, dass die Erfahrung, in der Autonomie hinsichtlich der Prüfungen beschränkt worden zu sein, in Beziehung zu einer Schwächung der kollektiven Selbstwirksamkeit steht.

Hinsichtlich persönlicher Faktoren zeigt die Geschlechtszugehörigkeit keinen Zusammenhang mit der wahrgenommenen kollektiven Selbstwirksamkeit.

Hinsichtlich der Dienstdauer und der kollektiven Selbstwirksamkeitserwartung ist ein U-förmiger Zusammenhang zu beobachten: Lehrpersonen, die weniger als sechs Jahre (standardisierter Koeffizient: 0.08), und Lehrpersonen, die mehr als 21 Jahre an einer Schule arbeiten (standardisierter Koeffizient: 0.05), halten das Kollegium für selbstwirksamer als ihre Kolleg/-innen, die 6 bis 20 Jahre unterrichten. Hinsichtlich des Bundeslandes zeigt sich, dass die hessischen Lehrer/-innen die kollektive Selbstwirksamkeit höher einschätzen als ihre Bremer Kolleg/-innen (standardisierter Koeffizient: -0.08). Weitere Analysen müssten untersuchen, welche Faktoren für diese unterschiedlichen Entwicklungen in den Bundesländern ausschlaggebend sind.

### 5.5 *Fazit*

Wichtige Elemente der Schulqualität sind u.a. das Schulklima, die Arbeitszufriedenheit und die individuelle sowie die kollektive Selbstwirksamkeit der Lehrpersonen. Die Ergebnisse dieser Studie zeigen, dass die Voraussetzungen für das Erreichen einer wünschenswerten Schulqualität und damit verbunden einer wünschenswerten Leistung des Bildungssystems hinsichtlich des Schulklimas und der Arbeitszufriedenheit ganz gut erfüllt sind.

Der Großteil der Lehrpersonen nimmt das Schulklima eher positiv wahr und ist mit der Arbeit eher zufrieden bis zufrieden. Die Schüler/-innen schätzen das Schulklima als eher positiv ein, aber ihre Einschätzung divergiert von denen der Lehrpersonen. Hier wäre weitere Forschung interessant. Zudem wurde in dieser Studie zwar die Schulebene berücksichtigt, 8% der Varianz lässt sich hier erklären, Aspekte wie die Größe und Lage der Schule oder die Zusammensetzung der Schüler/-innenschaft konnten jedoch nicht berücksichtigt werden. Schließlich handelt es sich bei den untersuchten Schulen – aufgrund des Schwerpunkts Zentralabitur – einzig um Gymnasien. Wie das Schulklima an Real-, Haupt- und Gesamtschulen beschaffen ist, konnte hier nicht analysiert werden.

Auch hinsichtlich der hohen Zustimmungsraten zur Arbeitszufriedenheit muss das bereits angesprochene Problem berücksichtigt werden, dass Arbeitszufriedenheit subjektiv hergestellt wird und abhängig ist von dem Anspruchsniveau, der Problemwahrnehmung und der Situationsbewältigung (Bruggemann, Groskurth & Ulich, 1975, S. 132). Arbeitszufriedenheit, wie sie hier gemessen wurde, bezieht somit die Bewertung der eigenen Arbeit mit ein. Um objektivere Kriterien für die Zufriedenheit mit den Arbeitsbedingungen zu erhalten, müssten z.B. Fehlzeiten und Fluktuation der Mitarbeitenden ausgewertet und Belastungen am Arbeitsplatz erhoben werden (Grunder & Bieri, 1995, S. 63; Merz, 1979a, zitiert nach Ammann, 2004, S. 60).

In Bezug auf die individuelle und kollektive Selbstwirksamkeit, die den Umgang mit Herausforderungen für den Einzelnen und die Gruppe erleichtern können und damit auch die Umsetzung von Reformen, ließen sich in dieser Untersuchung teilweise verbesserungswürdige Resultate beobachten. 9% bis 22% der Lehrpersonen erleben sich selbst als eher nicht selbstwirksam und fast die Hälfte der Lehrpersonen nimmt das Kollegium als eher nicht selbstwirksam wahr. Auch wenn über die drei Jahre in diesen beiden Aspekten eine leichte durchschnittliche Zunahme der Selbstwirksamkeit beobachtet werden kann, bleibt eine relativ große Gruppe von Lehrpersonen mit einer eher geringen Selbstwirksamkeitsüberzeugung hinsichtlich individueller und kollektiver Handlungsfähigkeit. Hinsichtlich dieser beiden Aspekte wären demnach Maßnahmen und Projekte hilfreich, die die Selbstwirksamkeitserwartungen der Lehrer/-innen steigern.

Die Erfahrung mit den zentralen Prüfungen gehört nicht zu solchen Maßnahmen, denn hier zeigt sich ein leichter, signifikanter, negativer Zusammenhang. Dieser war zu erwarten, da das Instrument Zentralabitur in die Autonomie der Lehrpersonen eingreift, was als Zuschreibung von Inkompetenz in diesem Bereich erlebt werden könnte. Zudem wird durch die eingeführte Unsicherheit gegenüber den Prüfungsfragen im Abitur, den Lehrpersonen auch tatsächlich Handlungs- und Planungssicherheit genommen. Beides kann sich negativ auf die Wahrnehmung der Selbstwirksamkeit auswirken.

Auch hinsichtlich der Arbeitsunzufriedenheit ist, wie erwartet, ein schwacher Unzufriedenheitsanstieg im Zusammenhang mit der Erfahrung mit den zentralen Prüfungen zu beobachten. Dies mag daran liegen, dass das Zentralabitur erst einmal Umstellungen erfordert, die als Belastung erlebt werden können. Zudemengt es Handlungsspielräume der Lehrpersonen ein, was ebenfalls als negativ empfunden werden kann. Schließlich könnte dies auch damit zusammenhängen, dass die Lehrpersonen ihren Status in der Gesellschaft als niedriger einschätzen als zu Zeiten des dezentralen Prüfungssystems und daher auch unzufriedener mit ihrer Arbeit sind. Denn Bishop (1999, S. 389) stellte in einem internationalen Vergleich fest, dass Lehrpersonen in Ländern mit zentralen Prüfungen ihre eigene gesellschaftliche Position niedriger bewerten als Kolleg/-innen in Ländern ohne zentrales Prüfungssystem. Sie seien zudem signifikant eher bereit, z.B. die Arbeitsstelle zu wechseln.

Angesichts der zentralen Prüfungen bleibt lediglich die Wahrnehmung des Schulklimas konstant. Dies entspricht der Erwartung, dass das hier hauptsächlich gemessene soziale Klima an der Schule von Änderungen des Prüfungssystems unberührt bleibt – besonders wo diese Änderungen außerhalb der Schule initiiert wurden.

Schließlich muss angemerkt werden, dass alle untersuchten Faktoren nur geringe Schwankungen in den Konstrukten erklären. Dies liegt hinsichtlich des Schulklimas und der Arbeitsunzufriedenheit sicherlich daran, dass insgesamt auch wenig Varianz aufgeklärt werden kann. In Bezug auf die Selbstwirksamkeitserwartungen jedoch liegt die Ursache auch darin, dass relevante Faktoren für die Entwicklung von Selbstwirksamkeit im Rahmen dieser Studie nicht berücksichtigt werden konnten (z.B. Erfahrung von Selbstwirksamkeit im täglichen Arbeitskontext, Verhalten der Schulleitung, etc.).

## 6 Literatur

- Ammann, T. (2004). *Zur Berufszufriedenheit von Lehrerinnen: Erfahrungsbilanzen in der mittleren Berufsphase*. Bad Heilbrunn/Obb.: Klinkhardt.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28 (2), 117-148.
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York, N.Y.: Freeman.
- Bauer, W. (1980). *Lehrerarbeit und Lehrerbewußtsein: Empirische Untersuchung zum Verhältnis von Arbeits- und Berufssituation der Lehrer und Lehrerbewußtsein*. Minerva-Fachserie Pädagogik. München: Minerva-Publikation.
- Berry, W.D. & Feldman, S. (1985). *Multiple regression in practice*. Beverly Hills/London/New Delhi: Sage.
- Berryhill, J., Linney, J.A. & Fromewick, J. (2009). The effects of education accountability on teachers: are policies too-stress provoking for their own good? *International Journal of Education Policy and Leadership*, 4 (5), 1-14.
- Bessoth, R. & Weibel, W. (2003). *Führungsqualität an Schweizer Schulen: Werkzeuge zu Klima, Kultur und zur Kompetenz der Führenden* (2., überarb. und erw. Aufl.). Aarau: Sauerländer.
- Bishop, J.H. (1999). Are national exit examinations important for educational efficiency? *Swedish Economic Policy Review*, 6, 349-398.
- Bohrnstedt, G.W. & Carter, T.M. (1971). Robustness in regression analysis. In H.L. Costner (Hrsg.), *Sociological Methodology* (pp. 118-146). San Francisco: Jossey-Bass.
- Brockmeyer, R. & Edelstein, W. (1997). *Selbstwirksame Schulen: Wege pädagogischer Innovation*. Oberhausen: Laufen.
- Bruggemann, A., Groskurth, P. & Ulich, E. (1975). *Arbeitszufriedenheit*. Bern: Huber.
- Buholzer, A. (2000). *Das Innovationsklima in Schulen*. Aarau: Bildung Sauerländer.
- Dick, R. van (2006). *Stress und Arbeitszufriedenheit bei Lehrerinnen und Lehrern: Zwischen „Horrorjob“ und Erfüllung* (2., leicht veränd. Aufl.). Marburg: Tectum.
- Eder, F. (1994). *Schul- und Klassenklima: Ausprägung, Determinanten und Wirkungen des Klimas an weiterführenden höheren Schulen*. Innsbruck: Österreichischer Studienverlag.
- Eder, F. (1996). *Schul- und Klassenklima: Ausprägung, Determinanten und Wirkungen des Klimas an höheren Schulen*. Innsbruck: Studien-Verlag.
- Eder, F. (1998). *Linzer Fragebogen zum Schul- und Klassenklima für die 8.-13.Klasse (LFSK 8-13)*. Göttingen: Hogrefe.
- Fend, H. (1977). *Schulklima: soziale Einflussprozesse in der Schule*. Forschungsarbeiten aus dem Zentrum I Bildungsforschung, Sonderforschungsbereich 23 der Universität Konstanz (Bd. 3). Weinheim: Beltz.
- Fend, H. (1986). „Gute Schulen – schlechte Schulen“. Die einzelne Schule als pädagogische Handlungseinheit. *Die Deutsche Schule*, 78 (3), 275-293.

- Fend, H. (1998). *Qualität im Bildungswesen: Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. Weinheim: Juventa.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen: Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Grunder, H.-U. & Bieri, T. (1995). *Zufrieden in der Schule? Zufrieden mit der Schule? Berufszufriedenheit und Kündigungsgründe von Lehrkräften*. Bern: Haupt.
- Halbheer, U. & Kunz, A. (2011). *Kooperation von Lehrpersonen an Gymnasien: Eine qualitative und quantitative Analyse der Wahrnehmung von Lehrpersonen aus schul- und governance-theoretischer Perspektive*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Halbheer, U., Kunz, A. & Maag Merki, K. (2005). *Pädagogische Entwicklungsbilanzen an Zürcher Mittelschulen. Indikatoren zu Kontextmerkmalen gymnasialer Bildung. Perspektive der Lehrpersonen: Schul- und Unterrichtserfahrungen. Skalen- und Itemdokumentation*. Zürich: Forschungsbereich Schulqualität & Schulentwicklung, Pädagogisches Institut, Universität Zürich.
- Helsper, W., Böhme, J., Kramer, R.T. & Lingkost, A. (2001). *Schulkultur und Schulmythos: Gymnasien zwischen elitärer Bildung und höherer Volksschule im Transformationsprozess*. Opladen: Leske + Budrich.
- Hox, J. (2002). *Multilevel analysis – techniques and applications*. Mahwah: Erlbaum.
- Jonas, K. & Brömer, P. (2002). Die sozial-kognitive Theorie von Bandura. In D. Frey & M. Irle (Hrsg.), *Theorien der Sozialpsychologie* (4., vollst. überarb. und erw. Aufl.; S. 277–299). Heidelberg: Springer.
- Knippenberg, D. van, Knippenberg, B. van, Cremer, D. de & Hogg, M. (2004). Leadership, self, and identity: a review and research agenda. *The Leadership Quarterly*, 15 (6), 825–856.
- Kolbe, M., Jerusalem, M. & Mittag, W. (1998). Veränderung von Selbstwirksamkeit und Klassenklima im zeitlichen Verlauf. *Unterrichtswissenschaft*, 26 (2), 116–127.
- Leutwyler, B. & Maag Merki, K. (2005). *Mittelschülerhebung 2004. Indikatoren zu Kontextmerkmalen gymnasialer Bildung. Perspektive der Schülerinnen und Schüler: Schul- und Unterrichtserfahrungen. Skalen- und Itemdokumentation*. Zürich: Forschungsbereich Schulqualität & Schulentwicklung, Pädagogisches Institut, Universität Zürich.
- Litwin, G.H. & Stringer, R.A. (1968). *Motivation and organizational climate*. Boston: Harvard University Press.
- Merz, J. (1979a). *Berufszufriedenheit von Lehrern: Eine empirische Untersuchung*. Weinheim: Beltz.
- Merz, J. (1979b). Wo drückt der Schuh am meisten? Empirische Befunde zur subjektiven Berufsbelastung von Lehrern. *Bayerische Schule*, 32, 13–16.
- Merz, J. (1980). Eine Skala zur Messung der allgemeinen Berufszufriedenheit von Lehrern (ABZ). *Zeitschrift für Empirische Pädagogik*, 4, 59–69.
- Opdenakker, M.-C. & Damme, J. van (2006). Teacher characteristics and teaching styles as effectiveness enhancing factors of classroom practice. *Teaching and teacher education*, 22, 1–21.
- Oswald, F., Pfeifer, B., Ritter-Berlach, G. & Tanzer, N. (1989). *Schulklima: Die Wirkungen der persönlichen Beziehungen in der Schule*. Wien: Universitätsverlag.
- Payne, R. & Pugh, D.S. (1983). Organizational structure and climate. In M.D. Dunnette, L.M. Hough & H.C. Triandis (Eds.), *Handbook of industrial and organizational psychology* (pp. 1125–1173). New York: Wiley.
- Raudenbush, S.W., Bryk, A.S. & Congdon, R. (2004). *HLM 6 for Windows [Computer software]*. Lincolnwood, IL: Scientific Software International, Inc.
- Rutter, M., Maughan, B., Mortimore, P. & Ouston, J. (1980). *Fünfehtausend Stunden: Schulen und ihre Wirkung auf die Kinder*. Weinheim: Beltz.
- Satow, L. & Bäßler, J. (1998). Selbstwirksamkeit und körperliches Befinden Jugendlicher. *Unterrichtswissenschaft*, 26 (2), 127–140.
- Schmitz, G.S. (1998). Entwicklung der Selbstwirksamkeitserwartungen von Lehrern. *Unterrichtswissenschaft*, 26 (2), 140–158.

- Schmitz, G.S. (2001). Kann Selbstwirksamkeitserwartung Lehrer vor Burnout schützen? Eine Längsschnittstudie in zehn Bundesländern. *Psychologie in Erziehung und Unterricht*, 48 (1), 49–67.
- Schmitz, G.S. & Schwarzer, R. (2000). Selbstwirksamkeitserwartung von Lehrern: Längsschnittbefunde mit einem neuen Instrument. *Zeitschrift für Pädagogische Psychologie*, 14 (1), 12–25.
- Schwarzer, R. & Jerusalem, M. (Hrsg.) (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin und Humboldt-Universität Berlin.
- Schwarzer, R. & Schmitz, G.S. (1999). Kollektive Selbstwirksamkeitserwartung von Lehrern: Eine Längsschnittstudie in zehn Bundesländern. *Zeitschrift für Sozialpsychologie*, 30 (4), 262–274.
- Six, B. & Kleinbeck, U. (1989). Arbeitsmotivation und Arbeitszufriedenheit. In E. Roth, H. Schuler & A.B. Weinert (Hrsg.), *Organisationspsychologie* (Wirtschafts-, Organisations- und Arbeitspsychologie; S. 348–398). Göttingen: Hogrefe.
- Spieß, E. & Rosenstiel, L. von (2010). *Organisationspsychologie: Basiswissen, Konzepte und Anwendungsfelder*. München: Oldenbourg.
- Spieß, E. & Winterstein, H. (1999). *Verhalten in Organisationen: Eine Einführung*. Stuttgart: Kohlhammer.
- Thomas, A.R. (1976). The organizational climate of schools. *Internationale Zeitschrift für Erziehungswissenschaft*, 22 (4), 440–463.
- Ulich, K. (1998). Schulische Sozialisation. In K. Hurrelmann & D. Ulich (Hrsg.), *Handbuch der Sozialisationsforschung* (5., neu ausgest. Aufl., Studienausg.; S. 377–396). Weinheim: Beltz.
- Varbelow, D. (2003). *Schulklima und Schulqualität im Kontext abweichender Verhaltensweisen*. Marburg: Tectum-Verlag.
- Zhang, D. (2005). *A Monte Carlo investigation of robustness to nonnormal incomplete data of multi-level modelling*. Texas A&M University. Accessible: <http://repository.tamu.edu/bitstream/handle/1969.1/4405/etd-tamu-2005B-EPSY-Zhang.pdf?sequence=1> [27.12.2010].



Daniela J. Jäger (2014): Zwischen Empowerment und Kontrolle. Die praktische Umsetzung des New Public Management und Professioneller Lerngemeinschaften in Kanada. Eine Fallstudie. *Zeitschrift für Pädagogik*, 60 (2), 303-324.

Hinweis

Sperrfrist bis 01.05.2015.

## **Zwischen Empowerment und Kontrolle – Steuerung in Alberta, Kanada**

Eine Analyse der Umsetzung von NPM und PLG in einem „best-case“ der Bildungsperformanz

*Daniela J. Jäger*

### *Zusammenfassung*

*Um zu untersuchen, welche Merkmalskombinationen eine positive Bildungsperformanz erzeugen, wird in diesem Artikel anhand einer best-case Fallanalyse die Umsetzung von New Public Management (NPM) in Kombination mit Professionellen Lerngemeinschaften (PLG) in Alberta, Kanada, analysiert. Mittels einer Daten- und Methodentriangulation werden die spezifischen Merkmalskombinationen sowie Unterstützungssysteme herausgearbeitet. Zudem wird die Kombination von NPM und PLG diskutiert. Es zeigt sich u.a., dass Wettbewerbselemente des NPM eine geringere Rolle spielen als die Zusammenarbeit der Akteuer auf horizontaler und vertikaler Ebene (PLG), die Professionalisierung sowie die Unterstützung der Akteure.*

*Keywords: Governance, Kanada, New Public Management, Professionelle Lerngemeinschaften, Unterstützungssysteme*

Daniela J. Jäger

Malzstr.1

8045 Zürich

061-5356099

daniela-julia.jaeger@gmx.net

Universität Zürich

Institut für Erziehungswissenschaft

Freiestrasse 36

8032 Zürich

Telefon: +41 44 634 45 70

Raum: FRE G20

djaeger@ife.uzh.ch

## **Zwischen Empowerment und Kontrolle – Steuerung in Alberta, Kanada**

Eine Analyse der Umsetzung von NPM und PLG in einem „best-case“ der Bildungsperformanz

### *Zusammenfassung*

*Um zu untersuchen, welche Merkmalskombinationen eine positive Bildungsperformanz erzeugen, wird in diesem Artikel anhand einer best-case Fallanalyse die Umsetzung von New Public Management (NPM) in Kombination mit Professionellen Lerngemeinschaften (PLG) in Alberta, Kanada, analysiert. Mittels einer Daten- und Methodentriangulation werden die spezifischen Merkmalskombinationen sowie Unterstützungssysteme herausgearbeitet. Zudem wird die Kombination von NPM und PLG diskutiert. Es zeigt sich u.a., dass Wettbewerbselemente des NPM eine geringere Rolle spielen als die Zusammenarbeit der Akteure auf horizontaler und vertikaler Ebene (PLG), die Professionalisierung sowie die Unterstützung der Akteure.*

## **1 Einleitung**

Viele Staaten haben Instrumente des New Public Management (NPM) zur Steuerung ihrer Bildungssysteme eingeführt (vgl. OECD, 2004). Freie Schulwahl, leistungsabhängige Bezahlung, zentrale Prüfungen und Schulautonomie sollen eine Effizienzsteigerung ermöglichen (vgl. Schedler & Proeller, 2006). Dabei erzeugt NPM im Bildungswesen laut empirischen Studien eine Reihe negativer Effekte (vgl. für einen Überblick Bellmann & Weiß) und scheint zudem nicht dem bürokratischen Modell, der hierarchischen Inputsteuerung, überlegen zu sein (vgl. Altrichter & Helm, 2011). Altrichter und Helm (2011, S. 31) resümieren daher, dass es nicht die Megafaktoren seien, die einen Einfluss haben, sondern eher die spezifischen Konstellationen. Auch Thillmann (2011, S. 75) hebt hervor, dass es notwendig sei, steuerungstheoretische Informationen über „effektive Merkmalskombinationen“ zu identifizieren. Zudem werden in der Literatur immer wieder Unterstützungssystemen als Lösungsansatz für die Verminderung von Ungleichheiten im Bildungssystem genannt (z.B. DIPF, 2007; van Ackeren, 2010) – ohne dass diese im Detail erfasst und analysiert seien.

Um etwas über Merkmalskombinationen und spezifische Unterstützungssysteme zu erfahren, bieten sich Fallstudien an. In diesem Artikel werden diese daher anhand eines „best-case“ der Bildungsperformanz analysiert. Dazu wurde eine Schule in der kanadischen Provinz Alberta ausgewählt. Alberta erzielte nicht nur hervorragende Ergebnisse hinsichtlich der Bildungsperformanz (PISA-Ergebnisse vgl. OECD, 2007; OECD, 2010a), sondern in Alberta werden zwei Konzepte in Kombination umgesetzt, die weltweit diskutiert und (in Varianten) umgesetzt werden: NPM (Anfang der 90er Jahre) und Professionelle Lerngemeinschaften (PLG, seit 2004) (vgl. Evans, 1997; Bensen & Rolff, 2006; OECD, 2004; Rolff, 2001; Tarnoczi, 2006). Alberta kann damit auf eine relativ lange Tradition der PLG zurückblicken, wenn man bedenkt, dass Bensen und Rolff 2006 von den PLG als Novität im deutschsprachigen Raum berichten (Bensen & Rolff, 2006).

So stellen sich die Forschungsfragen: Welche Elemente des NPM und der PLG werden umgesetzt? Welche Probleme ergeben sich aus der Kombination der beiden Konzepte? Und: Lassen sich Hypothesen aus dem gewonnenen Steuerungswissen ableiten? Die Ergebnisse basieren auf quantitativ und qualitativ gewonnenen Daten, die 2010 mit Unterstützung der Stiftung für Kanada-Studien und des Deutschen Akademische Austauschdienst erhoben wurden.

## **2 New Public Management und Professional Learning Communities**

Unter New Public Management fallen Reformen der öffentlichen Verwaltung, die Output-basierte Instrumente der Steuerung beinhalten (Schedler & Proeller, 2006). Das Output, z.B. Bildungsziele, wird vom Prinzipal, z.B. der Regierung, definiert (zentrale Steuerung) und die Agenten, z.B. die Schulleitung oder die LehrerInnen, sollen diese Ziele umsetzen. Entscheidend dabei ist, dass der Agent Freiräume hinsichtlich des Weges zur Zielerreichung haben soll, um so sein lokales Wissen nutzen zu können. Im besten Fall sollen sich alle betroffenen Akteure an Entscheidungen beteiligen. Um dies zu ermöglichen, sollen wichtige Kompetenzen wie Personal- und Budgetfragen auf die Schulebene verlagert werden (dezentrale Elemente) (vgl. u.a. Schedler & Proeller, 2006). Auch Fragen zur Unterrichtsgestaltung sollen z.B. an die Bedürfnisse der SchülerInnen angepasst werden können. Strategische Entscheidungen werden somit von der zentralen Ebene, operative von der dezentralen getroffen. Dies impliziert eine Trennung der operativen und der strategischen Ebene und ermöglicht es zudem, Wettbewerb zwischen den Agenten einzuführen.

Der Wettbewerb soll effektive und effiziente Lösungen im Sinne des Prinzipals und der „Kunden“ hervorbringen (von Hayek, 1969). So kann zum einen der Prinzipal das Verhalten der Agenten sanktionieren (Boni, Kürzungen, Kündigung). Zum zweiten sollen die Agenten auch um die Gunst der Kunden werben (Tiebout, 1956). Zwei daraus folgende, zentrale Aspekte im Bildungswesen sind die Forderung nach freier Schulwahl und nach öffentlich vergleichbaren Daten zu den Schulen (Rankings). Eltern sollen aufgrund der in Rankings bereitgestellten Information entscheiden, zu welcher Schule sie ihr Kind schicken. Je nach Anmeldungen an den Schulen müssen diese verkleinert, geschlossen oder ausgebaut werden. Die Performanz des Systems soll durch diese Selektion verbessert werden. In Tabelle 1 sind die Instrumente des NPM zusammengefasst.

Professional Learning Communities (PLC) wurden im deutschsprachigen Raum als Professionelle Lerngemeinschaften (PLG) von Bonsen und Rolff (2006) eingeführt. Die Begriffe werden im Folgenden synonym verwendet, auch wenn es geringe Abweichungen im Verständnis gibt (vgl. Fussangel, 2008).

PLG beruhen auf der Annahme situierten Lernens (Alberta Education, 2006, S. 5; Fussangel, 2008, S. 93-96). Situiertes Lernen bedeutet, dass eine Ähnlichkeit zwischen der Situation, in der gelernt wurde und der, in der das Wissen angewendet werden soll, Lernen wie Umsetzung erleichtert (Fussangel, 2008, S. 94). PLG sind daher im Arbeitskontext, z.B. in Schulen, angesiedelt. Die gemeinsame Reflektion soll Lernprozesse bei den Teilnehmenden anstoßen, die maßgeschneidert für die jeweilige Situation sind (Geringer, 2003, S. 373-380, zit. nach Alberta Education, 2006, S. 6). Dies soll es erleichtern, neue Erkenntnisse in Handlungen umzusetzen (vgl. Fussangel, 2008) und so Lernfortschritte bei den SchülerInnen bewirken (vgl. Alberta Education, 2006).

Im Idealfall wird sich erhofft, dass die PLGs einen Schritt Richtung Lernender Organisation sind (Bonsen & Rolff, 2006, S. 182). So schließt Morrissey (2000, S. 10): „a professional learning community becomes the supporting structure for schools to continuously transform themselves through their own internal capacity“ und Tarnoczy (2006) beschreibt die PLG als ein „elastic set of organizational structures and practices“, da sie von der Idee des ständig fortschreitenden Wachstums getragen sind und dies eine stete Anpassung, das heißt, Veränderung der Organisation impliziert.

Eine gelungene Zusammenarbeit in den PLG soll durch die Verwirklichung von Kollegialität und Professionalisierung gelingen (Marzano, 2003, S. 6, zit. nach Alberta

Education, 2006). Unter Kollegialität fasst Marzano (2007, S. 603) geteilte Normen, Zusammenarbeit und Mitbestimmungsrechte an Entscheidungen. Professionalisierung zeichnet sich nach ihm durch Weiterbildungsaktivitäten aus. Alberta Education (2006, S. 4) nennt zudem die Unterstützung der Akteure (mit Rückgriff auf Morrissey, 2000, S. 10) als relevantes Charakteristikum für PLGs. Unterstützung bedeutet bei Morrissey (2000), ein adäquater struktureller Rahmen, Zeit für Kooperation und Kommunikationsprozesse, und ein vertrauensvoller Umgang zwischen den Akteuren (vgl. auch Hord, 1997). Um die Qualität der Arbeit in den PLGs zu beschreiben, werden in der Literatur noch weitere Faktoren genannte, auf die hier aus Platzgründen nicht eingegangen werden kann (vgl. z.B. Bonsen & Rolff, 2006).

Tabelle 1: Charakteristika von NPM und PLG

NPM	PLG
Freie Schulwahl	Kollegialität (geteilte Normen, Zusammenarbeit, Mitbestimmung)
Rankings und Schulbeurteilungen	Professionalisierung (Weiterbildung)
Sanktionierbarkeit der Akteure seitens der Regierung	Unterstützung (strukturell und sozial)
<b>Elemente der Zentralisierung</b>	
Definition von Bildungszielen seitens der Regierung	
Überprüfung von Bildungszielen seitens der Regierung	
<b>Elemente der Dezentralisierung</b>	
Personalautonomie der Schule	
Sachmittelbewirtschaftung seitens der Schule	
Lehrmittel sowie Unterrichtsgestaltung obliegen der Schule	
Beteiligung der Akteure der Schule an Entscheidungen	

## 2.1 Die Kombination von NPM und PLG – Gemeinsamkeiten und Widersprüche

Die Konzepte von NPM und PLG sind hinsichtlich der Kompetenzverschiebung von der Verwaltung zu den Schulen im Sinne der Dezentralisierung kompatibel. In den PLG benötigen die Akteure Entscheidungsspielräume, um eigene Ideen entwickeln und umsetzen zu können (vgl. Hord, 1997). Im NPM soll ebenfalls die Partizipation aller Beteiligten gewährleistet werden, um so das lokale Wissen für eine Anpassung des Angebots an den Bedarf nutzen zu können. Eine Kombination der Konzepte könnte damit die partizipativen Elemente in der Steuerung erhöhen.

Die zentrale Steuerung der Ziele des Bildungssystems (als Teil der NPM-Charakteristika) und die Forderung nach „geteilter Führung“ sowie nach einem System, welches auf allen Ebenen offen für Lernprozesse ist (vgl. Konzept der PLG) widersprechen sich dagegen im Grundsatz. Die Steuerungsmodi im NPM (Weisung,

vgl. Thillmann, 2011, S. 66) und in den PLG (Moderation und Kooperation, vgl. Thillmann, 2011, S. 66) lassen sich nur kombinieren, wenn man – wie im NPM – davon ausgeht, dass es möglich ist, die operative Ebene von der strategischen zu trennen. Wie Tarnoczi (2006) betont, ist es fraglich, inwiefern eine solche Trennung der operativen und der strategischen Ebene praktisch möglich ist. Sollen die Akteure in den PLG über Autonomie verfügen und zudem Lerneffekte auf allen Ebenen des Systems stattfinden (vgl. Bensen & Rolff, 2006, S. 182; Morrissey, 2000, S. 10), ist hier ein Spannungsverhältnis zu erwarten.

Ein weiteres Spannungsverhältnis ergibt sich hinsichtlich der Input- bzw. Outputsteuerung: NPM betont die Outputsteuerung und den Wettbewerb. Die PLG dagegen werden als unterstützende Struktur verstanden, die sich durch Vertrauen, Kommunikation und Zusammenarbeit auszeichnet. Diese Infrastruktur soll von der Schulverwaltung durch die Eingabe von Ressourcen ermöglicht werden (Hord, 1997), was einer Inputsteuerung nahe kommt.

Schließlich kollidiert der Wunsch eines vertrauensvollen und kooperativen Umgangs der Akteure mit der Idee des Wettbewerbs und der Steuerung durch Sanktionen. Es ist daher spannend zu untersuchen, wie mit diesen konzeptuellen Widersprüchen in der Praxis umgegangen wird.

## 2.2 Empirische Befunde zur Wirkung von NPM und PLG

In der Literatur finden sich zahlreiche empirische Befunde zur Wirkung einzelner Elementen des NPM und nur wenige empirische Studien zu Effekten der PLG. Zu Interaktionseffekten dieser Instrumente liegen bisher keine Untersuchungen vor.

Die Ergebnisse zu Wirkungen von Elementen des NPM zeigen, dass durch die „freie Schulwahl“ sich oftmals die soziale Segregation verstärkt (u.a. Böhlmark & Lindahl, 2007; Tomlinson, 2000) und dass die Idee leistungsabhängiger Bezahlung in einigen Fällen zu nicht-intendierten Effekten wie dem teaching und grading to the test führt (vgl. für einen Überblick: Bellmann & Weiß, 2009). Die bisherigen Ergebnisse zu den Effekten von Autonomie sind zudem inkonsistent (Altrichter & Rürup, 2010). Nur einzelne empirische, international vergleichende Studien heben Vorteile in der SchülerInnenleistung des NPM-Modells, besonders der zentralen Prüfungen, hervor (u.a. Wößmann, 2008). Die Methoden dieser Studien sind umstritten (vgl. zusammenfassend Altrichter & Rürup, 2010).

Von der Arbeit der PLGs wird sich in erster Linie eine Erhöhung der SchülerInnenleistungen erhofft (vgl. z.B. Alberta Education 2006:4). Die Meta-Analyse von Lomos, Hofman und Bosker (2011) bestätigt diese Erwartung, wenn auch bisher nur wenige quantitative empirische Studien vorliegen (5 Studien).

Eine vergleichende Fallstudie von Achinstein (2002, S.450) aus den USA zeigt aber auch, dass in PLGs die Gefahr der Stagnation der Arbeit besteht. Servage (2008) berichtet aus der Praxis in Alberta, dass die meiste Arbeit in den PLGs sich auf das Studium des Curriculums und die Auswertung von Daten u.a. beziehe. Dies führe dazu, dass die Lehrpersonen hauptsächlich auf relativ kurzfristige Ziele fokussiert seien und eine Nachhaltigkeit oder gar Veränderung der Kultur kaum stattfände (Servage, 2008).

Die Literatur zeigt zudem, dass es schwer ist, die Bedingungen für PLGs erfolgreich zu implementieren (vgl. u.a. Dooner, Mandzuk, Clifton, 2008; Leclerc, Moreau, Dumouchel & Sallafranque- St.-Louis, 2012; Sackney, 2011). Insbesondere wird hier die herausfordernden Rolle der Schulleitung, die als unterstützende, führende und gleichzeitig Führung-teilende definiert und als besonders wichtig erachtet wird, diskutiert (vgl. u.a. Burger, Nadirova, Brandon, Garneau & Gonnet, 2011; Sackney, 2007; Thompson, Gregg & Niska, 2004; Marzano, 2007).

### **3 Fragestellungen und Annahmen**

Aufgrund der bisher unzureichenden Information über gelungene Kombinationen von Steuerungselementen und den im NPM überwiegend negativen Effekten, sind in diesem Artikel zwei Fragen zu klären: (1) Welche Elemente des NPM und der PLG werden in einem „best-case“ de facto umgesetzt? (2) Welche Probleme und Vorteile ergeben sich aus der Kombination der beiden Konzepte? Zudem wird anhand gängiger Indikatoren wie Bildungshintergrund der SchülerInnen explorativ analysiert, um den Kontext der Schule zu beschreiben. Um die erste Frage zu beantworten, wird anhand der Theorie zu NPM und PLG untersucht, welche der Elemente umgesetzt werden. Hier wird erwartet, dass Wettbewerbselemente, die in anderen Staaten negative Effekte erzeugen, eher in den Hintergrund treten, und dass stattdessen eher unterstützende Faktoren relevant sind, wie in der Literatur zu Steuerungsstrukturen vermutet wird (u.a. DIPF, 2007; Van Ackeren, 2010). Hinsichtlich der zweiten Frage zu den Problemen und Vorteilen der Kombination der Konzepte, wird explorativ ana-



lysiert, inwiefern die beschriebenen Spannungsverhältnisse zwischen Zentralisierung und Partizipation, Output- und Inputsteuerung sowie Wettbewerb und Kommunikation aufgelöst werden. Im Fazit wird geschaut, inwiefern sich Hypothesen aus dem gewonnenen Steuerungswissen ableiten lassen.

## **4 Design und Methoden**

Eine Fallstudie ermöglicht die Erfassung komplexer Zusammenhänge (Muno, 2009, S. 121). Um detailliertes Wissen zu erhalten, wurde eine Datentriangulation hinsichtlich der befragten Personen und eine Methodentriangulation vorgenommen (vgl. u.a. Flick, 2008). Die Erhebung erfolgte im Frühjahr 2010.

### *4.1 Fallauswahl*

Die Fallauswahl erfolgte in zwei Schritten: So wurde der Staat Kanada sowie die Provinz Alberta hinsichtlich der guten Leistung und geringer Disparitäten zwischen den SchülerInnen, die einen positiven Sonderfall im OECD-Vergleich darstellen, ausgewählt (vgl. OECD, 2007; OECD, 2010a). Als Grundlage dienten die Daten der OECD (2007) aus 2006. Für die Wahl der Schule innerhalb der Großstadt Edmonton wurde sich auf ExpertInnenaussagen aus dem Edmonton Public School Board gestützt: Die ExpertInnen wurden um die Wahl einer Schule gebeten, die als durchschnittlicher Fall hinsichtlich der Sozialindikatoren in Edmonton gilt. So kann untersucht werden, wie in einem best-case der Bildungsperformanz im Jahr 2010 an einer durchschnittlichen Schule innerhalb einer Großstadt hohe Leistungen erbracht werden.

### *4.2 Datengrundlage, Methode und Auswertung*

Um ein möglichst umfassendes Bild des Falles zu erlangen, wurde angestrebt, alle Akteure des Mehrebenensystems zu befragen. Innerhalb der Schule wurde der SchülerInnenjahrgang aus der Stufe 9 ( $n = 194$ ) – in Anlehnung an die PISA-Studie – sowie das gesamte Lehrkollegium ( $n = 30$ ) per Fragebogen befragt. Der Rücklauf bei den Lehrpersonen mit 73% ist sehr zufriedenstellend, der für die SchülerInnen mit 40% ausreichend. Des Weiteren wurden an der Schule neun Interviews u.a. mit der Schulleitung, FachkoordinatorInnen und Verantwortlichen einzelner Programme (Lehrpersonen mit spezieller Funktion) und eine teilnehmende Beobachtung bei ei-

nem Elterntreffen durchgeführt. Zudem wurden VertreterInnen des Bezirks, die für die Schule zuständig sind, interviewt. Dies sind der Assistant Superintendent und der/die SozialarbeiterIn und der/die SchulpsychologIn. Ein Gespräch mit dem „Trustee“ und dem Superintendenten konnte leider nicht realisiert werden. VertreterInnen des Research Support Services des Edmonton Public School Boards diskutierten dafür im Rahmen eines Vortrags (12.02.2010) über das Projekt und boten Einblick in ihre Arbeit. Tabelle 2 zeigt mittels welcher Methoden die Merkmale der Schule und die Elemente des NPM und der PLG erfasst wurden.

Tabelle 2: Eingesetzte Methoden je nach Untersuchungsbereich

Methode	Merkmale der Schule	NPM	PLG
Dokumentenanalyse	durchgeführt	durchgeführt	durchgeführt
Standardisierter Fragebogen	Lehrpersonen SchülerInnen	Lehrpersonen	Lehrpersonen
Quantitative Netzwerkanalyse	-	-	Lehrpersonen
Interview	Schulleitung Stellv. Schulleitung	Schulleitung Stellv. Schulleitung LP_F Sozialarbeiterin Assistant Superintendent	Schulleitung Stellv. Schulleitung LP_F Sozialarbeiterin Assistant Superintendent
Teilnehmende Beobachtung	-	Elterntreffen	-
ExpertInnendiskussion	-	Interessierte des Research Support Services (Edmonton Public School Board)	-

Anmerkung: stellv. = stellvertretend; LP\_F = Lehrperson in spezieller Funktion

Folgend sind Quellen zu den Fragen dokumentiert, die nicht selbst entwickelt wurden. Die Interviews wurden anhand eines Leitfadens durchgeführt, der einige Fragen aus dem TALIS Fragebogen für SchulleiterInnen (OECD, 2010b, S. 249) enthält. Im Ergebnisteil wird auf Aussagen aus den Interviews von der Schulleitung mit (SL) und von dem/ der Assistant Superintendent mit (AS) verwiesen. Die übrigen Akteure werden ausgeschrieben.

Folgend sind die Fragen aus dem standardisierten Fragebogen dargestellt, deren Ergebnisse hier berichtet werden.

- Anzahl der Bücher zu Hause (SchülerInnen): Einzelitem: „How many books do you have at home?“ Antwortoptionen: 0-10, 11-25, 26-100, 101-200, 201-500, mehr als 500. Quelle: OECD, 2006
- Schulklima (Lehrpersonen): Skala (5 Items): “Please indicate to what extent the left hand side or right hand side characteristics apply from your point of view.

Most of the time, the atmosphere at our school is.... happy – cheerless.” Quelle: In Anlehnung an Eder (1998, S. 12) (stark gekürzt)

- Geteilte Normen (Lehrpersonen): Zwei Einzelitems: „I am sure that we teachers can create a good atmosphere at our school on the basis of our joint actions, even if the workload is overwhelming.” Und: „Given that we pursue the same pedagogical intentions, we the teachers are able to cope with ‘difficult’ students at our school, too.” Antwortoptionen: strongly disagree (1), disagree (2), agree (3), strongly agree (4). Quelle: Maag Merki, Halbheer & Kunz (2005, S. 11), in Anlehnung an die Skala von Schwarzer und Jerusalem (1999)
- Strukturelle Unterstützung (Lehrpersonen): Zwei Einzelitems: „When the schedule was drafted, teamwork opportunities were considered.” Und “There is plenty of working room available for teamwork.” Antwortoptionen: „strongly disagree“ (1) bis „strongly agree“ (4). Quelle: Maag Merki, Halbheer & Kunz (2005, S. 6)
- Herausforderungen an der Schule: Einzelitem: SchülerInnen: „Please name the biggest problems at your school.” Lehrpersonen: “Please name the biggest challenges at your school.” Antwort: offen. Quelle: Eigenentwicklung
- Nutzung von Weiterbildungsangeboten (Lehrpersonen). Einzelitem: “How many one-day, 2-3 day, 3 day plus and/or ongoing or regular (e.g. weekly) trainings did you attend during the last 12 months?” Antwortoptionen: 0-1, 2-3, 4-5, 6 or more. Quelle: Eigenentwicklung
- Mitbestimmung der Lehrpersonen an der Schule (Lehrpersonen): Zwei Einzelitems: „The teaching staff has a lot of influence at our school.“ Und: „The teachers at our school participate in many decisions.” Antwortoptionen: „strongly disagree“ (1) bis „strongly agree“ (4). Quelle: Eigenentwicklung in Anlehnung an Partizipation der SchülerInnen bei Eder (1998)

#### *Quantitative Netzwerkanalyse*

- Kooperation: „With whom do you exchange expertise on social behaviour of students?“ Aufgrund der zuvor geführten Interviews wurde eine Akteursliste generiert, die den Lehrpersonen anschließend im Rahmen der Befragung vorgelegt wurde. Sie konnten angeben, ob sie mit den genannten Personen/ Institutionen kooperieren oder nicht (Vorgehen nach Jansen, 2006).

## 5 Merkmale und Kontext der untersuchten Schule

In Kanada ist Bildung Aufgabe der Provinzen (DIPF, 2007). Die Provinz ist in Bezirke („wards“) aufgeteilt, in denen VertreterInnen für das Politikfeld Bildung gewählt werden (Trustees) (EPSB, 2009a). Die Trustees sind im Bereich der Bildungspolitik Teil der Legislative (zusammen mit der Provinz, der sie unterstellt sind) und bilden das School Board. Die Spitze der Exekutive ist der Superintendent, er überwacht gemeinsam mit seinen Assistant Superintendents die Umsetzung der Gesetze<sup>1</sup>. Die kanadischen Schulen inkludieren nahezu alle SchülerInnen in den regulären Unterricht (Interview der Schulleitung: SL). Um dies zu gewährleisten, werden an den Schulen unterschiedliche Programme für die SchülerInnen angeboten und es stehen BeraterInnen zu bestimmten Themen zur Verfügung. Für letzteres kann die Schule beim Consulting Service des Distrikts die Leistung von LogopädInnen, SchulpsychologInnen, DrogenberaterInnen und weiteren ExpertInnen einkaufen (SL).

Die untersuchte Schule liegt urban in Edmonton, der Hauptstadt Albertas. Die Daten zur Leistung der SchülerInnen in den provinzwerten Tests zeigen, dass die SchülerInnenleistung im Durchschnitt Albertas und somit wahrscheinlich oberhalb des OECD-Durchschnitts liegt<sup>2</sup>. Die Streuung der Leistung der SchülerInnen aus Alberta im Lesen liegt im Jahr 2009 aber nur noch im Durchschnitt des OECD-Vergleichs (OECD, 2010a) (im Gegensatz zu der Datengrundlage von 2006 bei der Fallauswahl). Betrachtet man zudem die Schulabbrecherquote, so schneidet Alberta im innerkanadischen Vergleich unterdurchschnittlich ab, wobei sich eine Reduzierung der Abbrecherquote in den letzten Jahren zeigt (Schmold, 2010).

Das Kollegium an der Schule umfasst 30 Lehrpersonen (SL). Der familiäre Hintergrund der SchülerInnen ist eher bildungsnah (SL und Befragung der SchülerInnen). Spezielle Programme bietet die Schule für talentierte Kinder, für SchülerInnen mit weniger erfolgreichen Leistungen (zwei Programme) und für SchülerInnen, die kein oder nur wenig Englisch sprechen (SL). Letztere werden von einer speziell dafür ausgebildete Lehrperson unterrichtet (English Language Learning teacher (ELL)). Zudem wird SchülerInnen, etwa mit Autismus, ein „Teaching Assistant“ zur Seite gestellt (SL). Als externe Leistung hat die Schule ein/e SozialarbeiterIn

---

<sup>1</sup> [http://www.epsb.ca/policy/tr\\_sectionone\\_principles\\_of\\_board\\_oper.shtml](http://www.epsb.ca/policy/tr_sectionone_principles_of_board_oper.shtml), Zugriff: 23.05.2012

<sup>2</sup> Vgl. EPSB: <http://www.epsb.ca/about/testsrts.shtml>, Zugriff: 23.5.2012

(nach Bedarf) und ein/e SchulpsychologIn (ein bis zwei Mal im Monat) vom Distrikt engagiert (SL). Die SchülerInnen und Lehrpersonen nehmen als Probleme an ihrer Schule Konflikte des Miteinanders (Mobbing unter SchülerInnen) und der Schwierigkeitsgrad des Unterrichts wahr (standardisierte Befragung).

## 6 Wie wird NPM umgesetzt?

Im Folgenden wird dargelegt, inwiefern die in der Theorie dargestellten Elemente des NPM im untersuchten Fall umgesetzt werden.

Als Voraussetzung für NPM soll Wettbewerb durch freie Schulwahl und Rankings ermöglicht werden. Laut Gesetz herrscht in Alberta *freie Schulwahl*: So wirbt die Schulverwaltung: „Implemented more than two decades ago, Edmonton Public Schools created its revolutionary open boundary system“ (EPSB, 2010). Gleichzeitig gilt aber als erstes Auswahlkriterium der Einzugsbereich der Schule (SL und AS). Die wenigen freien Plätze würden per Losverfahren an die SchülerInnen verteilt, so der/die Assistant Superintendent. Die Schulleitung berichtet, dass sich die Schulleitungen vor einer Aufnahme einer SchülerIn absprechen. Es würden sonst oft SchülerInnen wechseln, die Problemen aus dem Weg gehen wollten, so die SL.

*Rankings*, die die durchschnittlichen Leistungen der SchülerInnen zwischen den Schulen vergleichen, wurden bis 2011 vor allem vom Fraser Institut publiziert (vgl. DIPF, 2007).<sup>3</sup> Laut Schulleitung lehnen die Schulen die Vergleiche stark ab, da Kontextfaktoren nicht berücksichtigt würden.

Insgesamt lässt sich festhalten, dass gesetzlich Wahlfreiheit gegeben ist, die Umsetzung aber schwierig ist und nur von wenigen SchülerInnen genutzt werden kann. Öffentlich zugängliche Rankings existierten zum Zeitpunkt der Erhebung (2010).

### 6.1 Elemente der Zentralisierung

Im NPM soll die zentrale Ebene Bildungsziele vorgeben und überprüfen sowie die Arbeit der Akteure überprüfen und sanktionieren.

---

<sup>3</sup> Seit 2011 wird dem Fraser Institut die Nutzung der Daten untersagt, da die Rangzahlen die Eltern in ihrer Entscheidung irreführen würden, vgl.

<http://education.alberta.ca/admin/resources/gla.aspx>; Zugriff: 02.01.13.

Im untersuchten Fall werden *Bildungsziele* z.B. durch das Curriculum von der Provinz vorgegeben (SL). Es werden provinzweite Tests während der Schulzeit in den Stufen 3, 6 und 9 und zum Abschluss der High School in Stufe 12 durchgeführt (Alberta Education, 2012). Die Tests sind eng an das Curriculum gebunden (DIPF, 2007, S. 273) und werden in den Kernfächern durchgeführt (Alberta Education, 2012).

Voraussetzung für Sanktionierung ist die Evaluation der Arbeit. Tabelle 3 zeigt im Überblick, *wessen Arbeit evaluiert bzw. sanktioniert* wird. Sanktionsmöglichkeiten können Boni und negative Sanktionen sein, aber auch die Befristung von Verträgen.

Es zeigt sich, dass die Arbeit eines jeden Verantwortlichen im Bildungssystem – über alle Hierarchien hinweg – evaluiert wird und zudem dass die Verträge fast aller Akteure – interessanterweise außer der Lehrpersonen (Provinz of Alberta, 2011; EPSB, 2009b) – befristet sind. Somit besteht theoretisch die Möglichkeit für Sanktionen.

Tabelle 3: Evaluierung und Sanktionierung der Akteure des Mehrebenensystems

	Evaluationen		Befristeter Vertrag	Sanktionen	
	Evaluation der Arbeit	Abstand der Evaluationen		Positiv (Boni)	Negativ
SchülerInnen	Noten	monatlich, jährlich, alle drei Jahre	-	(ja)	(ja)
Lehrpersonen	ja	jährlich	nein	nein	nein
Stellv. Schulleitung	ja	jährlich	ja: 1 Jahr	nein	nein
Schulleitung	ja	jährlich	Verhandlung	nein	nein
Schule	(ja)	sehr selten	-	nein	nein
Assistant Superintendent	ja	alle drei Jahre	ja: 3 Jahre	nein	nein
Superintendent	ja	wöchentlich, jährlich	ja: 5 Jahre	(ja)	(ja)
Trustee	ja	alle drei Jahre	ja: 3 Jahre	ja	ja

Anmerkung: stellv. = stellvertretend; Quellen: eigene Interviews mit den genannten Akteuren; DIPF, 2007; EPSB, 2009b; EPSB, 2010; Provinz of Alberta, 2011

Allerdings berichten alle befragten Akteure, dass bei einer Nicht-Erfüllung der vor der Evaluation festgelegten Ziele keine Sanktionen erfolgen, sondern Unterstützung. Die Leistung der SchülerInnen allerdings wird durch Noten und den Abschluss sanktioniert. Erbringen die SchülerInnen bei den Tests während der Schulzeit ungenügende Leistungen, so werden sie z.B. durch die Teilnahme an Förderprogrammen unterstützt (daher ist in Tabelle 3 das „Ja“ eingeklammert). Die Schule insgesamt wird nur sehr, sehr selten evaluiert, so die SL (daher eingeklammert). Der Superintendent kann bei ungenügenden Ergebnissen abgesetzt werden (Province of Alberta, 2011,

S. 88). Inwiefern dies geschieht, ist wegen des fehlenden Interviews nicht zu sagen (Klammer in Tabelle 3). Er erstattet den Trustees aber zusätzlich wöchentlich Bericht. Die Trustees werden durch die Wahl der BürgerInnen sanktioniert. Die BürgerInnen entscheiden, ob ein Trustee eine weitere Periode im Amt bleibt oder nicht.

Die Bildungsziele werden also von der Provinz vorgegeben und überprüft. Auch die Arbeit aller Akteure wird regelmäßig evaluiert, was die Möglichkeit von Sanktionen impliziert. Alle befragten Akteure berichten von unterstützenden Maßnahmen anstelle von Sanktionen im Falle einer negativen Evaluation. Die Trustees dagegen sowie die SchülerInnen erfahren Sanktionen.

## 6.2 Elemente der Dezentralisierung

Elemente der Dezentralisierung sind Autonomie auf Schulebene hinsichtlich der Personalauswahl, der Ressourcen, der Unterrichtsinhalte sowie eine Beteiligung der Akteure an Entscheidungsprozessen.

Bei *Personalentscheidungen* besteht nur eine sehr geringe Schulautonomie. Der Distrikt träge bei Neubesetzungen des Lehrpersonals eine Vorauswahl unter den BewerberInnen und böte der Schulleitung eine Liste von drei bis fünf KandidatInnen an (SL). Falls niemand auf die Stelle passe, könne die Liste aufgestockt werden, so die SL. Lehrpersonen kann nur bei deutlichen Verstößen vom Superintendent, nicht von der Schulleitung, gekündigt werden. Die Schulleitung kann entscheiden, welche Consultants sie einstellt.

Das *Budget* der Schule basiert hauptsächlich auf der Anzahl der SchülerInnen (SL, AS). Zusätzliche Gelder können die Schulen z.B. für SchülerInnen mit speziellem Förderbedarf erhalten oder aber, wenn sie spezielle neue Programme entwickeln oder testen (AS). Zudem werden Schulen, die trotz geringer Population erhalten bleiben sollen, subventioniert. Der Einsatz des Budgets der Schule kann, formal gesehen, von der Schule bestimmt werden, fließt aber hauptsächlich in die laufenden Personalkosten. Laut Gesetz dürfen die Personalkosten nicht 91% des Budgets überschreiten, so die SL. An der untersuchten Schule werden rund 91% in das Personal und die restlichen rund 9% in Weiterbildungen der Lehrpersonen und in die Ausstattung der Schule investiert. Neben den zugewiesenen Mitteln kann die Schule Gelder ansparen oder auch einwerben. Die Schulleitung berichtet, dass sie jedes

Jahr ein Plus erwirtschaftete, welches sich über die Jahre kumuliere und Flexibilität erlaube. Zu diesem Plus können zudem die Spenden der Eltern summiert werden.

Zur *Wahlfreiheit zu Unterrichtsinhalten* erläutert die Schulleitung den Einfluss der Provinz: Die Provinz definiere, welche Kurse verpflichtend sind und ob die School Boards zusätzliche Wahlfächer anbieten dürfen. Der Spielraum sinke von Ebene zu Ebene (Provinz, Distrikt, Schule, Lehrperson), so die Schulleitung, und: „Over the years that's been narrowed, narrowed, narrowed.“ Die Provinz autorisiere zudem die Schulbücher, die Liste sei oft sehr kurz: „Sometimes just one book“, so die Schulleitung.

Hinsichtlich der *Beteiligung der Akteure an Entscheidungen* erwarten die Trustees zumindest im Budgetierungsprozess eine Einbeziehung der schulischen Akteure (SL), die Umsetzung obläge aber der Schule. Den Lehrpersonen wird in einigen Bereichen Verantwortung übertragen, so die Schulleitung. Nach eigener Einschätzung haben die Lehrpersonen „großen“ (n = 22, 54,5%) bzw. sogar „sehr großen“ Einfluss (45,5%) und werden an vielen Entscheidungen beteiligt (n = 22; 59,1% stimmen zu, 40,9% stimmen sehr zu). Die Interessen der SchülerInnen sind, laut SL, nicht organisiert. Die Eltern treffen sich monatlich, sie haben aber kein formales Mitspracherecht (SL). Wenn der Budgetplan der Schule aufgestellt wird, werden sie von der Schulleitung eingeladen, ihn zu beurteilen (SL).

Insgesamt kann man sagen, dass hinsichtlich der dezentralen Elemente nur die Beteiligung der Akteure in Ansätzen (z.B. keine SchülerInnenbeteiligung) verwirklicht ist. Personalentscheidungen können nicht für die Lehrpersonen, also die größte Gruppe der Angestellten, getroffen werden, eine Auswahl der Beratenden an der Schule ist dagegen möglich. Von einer Lehrmittelfreiheit kann nicht gesprochen werden und inwiefern die Sachmittelbewirtschaftung als autonom gelten kann, ist fraglich.

## **7 Umsetzung der PLG?**

Im Folgenden wird erörtert, inwiefern die Vorraussetzung für eine Arbeit der PLG Kollegialität, Professionalisierung und Unterstützung im untersuchten Fall gegeben sind.



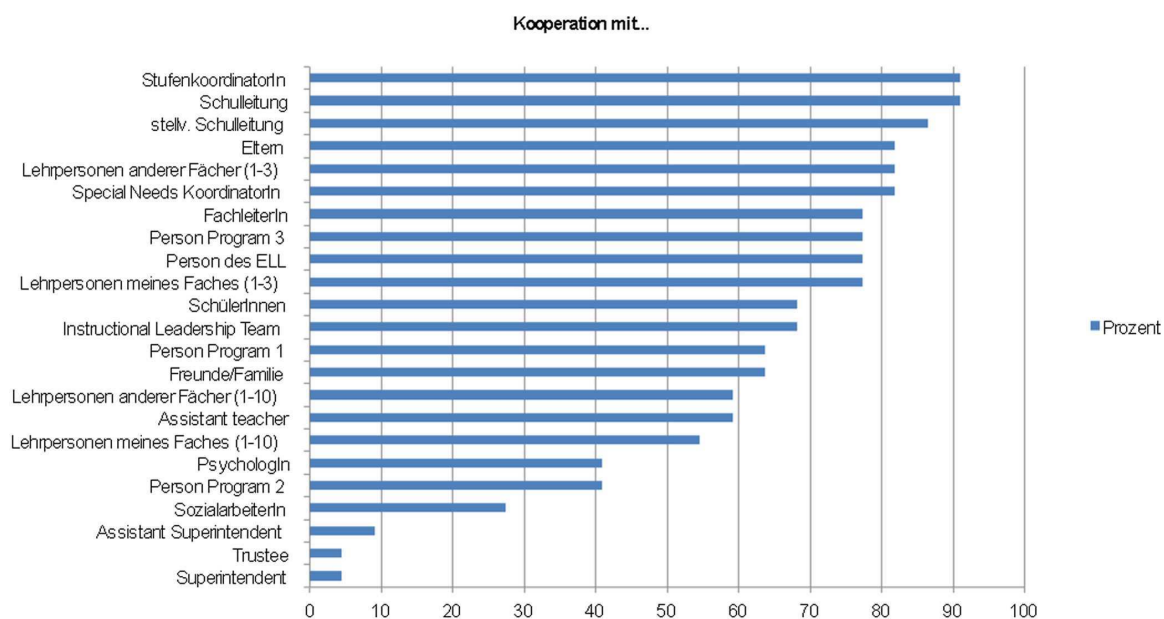
### 7.1 Kollegialität der Lehrpersonen

Kollegialität fasst Marzano (2007) als geteilte Normen, Zusammenarbeit und Mitbestimmung der Akteure. „Geteilte Normen“ werden hier anhand zweier Items konzeptualisiert. Die Zusammenarbeit wird über Angaben zur Atmosphäre an der Schule und die quantitative Netzwerkanalyse gemessen und die Mitbestimmungsmöglichkeiten wurden in der standardisierten Befragung der Lehrpersonen und im Interview mit der Schulleitung erfasst.

Im Durchschnitt stimmten die Lehrpersonen eher zu, dass sie aufgrund ihrer gemeinsamen Aktivitäten eine gute Atmosphäre an der Schule herstellen können, auch wenn der Arbeitsaufwand sehr hoch ist ( $n = 22$ ;  $M = 3,36$ ;  $SD = 0,49$ ). Die Lehrpersonen geben auch an, dass sie im Durchschnitt eher annehmen, auch unter schwierigen Umständen Ziele umzusetzen und mit „schwierigen SchülerInnen“ umgehen können ( $n = 22$ ;  $M = 3,18$ ;  $SD = 0,59$ ).

Wie Abbildung 1 zeigt, kooperieren die Lehrpersonen mit einer großen Anzahl Akteure (*Zusammenarbeit*).

Abbildung 1: Kooperation der Lehrpersonen



Anmerkungen: Erläuterungen zu den verschiedenen Positionen finden sich in Abschnitt 5 und 7.3. Die einzelnen Programme wurden aus Datenschutzgründen hier nicht entschlüsselt.

Die Lehrpersonen geben an, das Schulklima als positiv bis sehr positiv wahrzunehmen ( $n = 22$ ,  $M = 4,65$ ,  $SD = 0,43$ ). Die *Mitbestimmungsmöglichkeiten* sind aus Sicht der Lehrpersonen gut bis sehr gut verwirklicht, die SL gibt an, dass die Lehrpersonen in einigen Bereichen mitbestimmen könnten (vgl. Abschnitt 6.2).

Insgesamt scheint die Kollegialität unter den Lehrpersonen an der untersuchten Schule verwirklicht zu sein.

## 7.2 Professionalisierung

Die Professionalisierung wird über die Angebote an *Weiterbildung*, den Stellenwert dieser und die genutzte Weiterbildung seitens der Lehrpersonen gemessen.

Laut Schulleitung hält die Schule ein Budget für Weiterbildungen für die Lehrpersonen bereit. Die Umfrage unter den Lehrpersonen zeigt, dass mehr als ein Drittel der Lehrpersonen im letzten Schuljahr zwei bis drei eintägige Fortbildungen und jeweils 19% vier bis fünf bzw. mehr als sechs solcher Veranstaltungen besucht haben. 26% der Lehrpersonen geben sogar an, an mindestens zwei fortlaufenden Fortbildungen im letzten Jahr teilgenommen zu haben.

Der Distrikt bietet Weiterbildung für alle Akteursgruppen an, fördert aber auch den Besuch von Konferenzen, Masterstudiengängen oder eines Doktorats. Für die Fortbildung der (stellvertretenden) Schulleitungen gibt es auf der Distriktebene eine spezielle Fortbildungseinrichtung (Leading Service). Für die Personen des Consulting Service, z.B. die SozialarbeiterIn oder PsychologIn, werden zehn Mal im Jahr Weiterbildungstage angeboten. Diese sind für alle 150 BeraterInnen am gleichen Tag und gleichen Ort, um auch den informellen fachübergreifenden Austausch zu unterstützen (Aussage der SozialarbeiterIn). Weiterbildungen des Distrikts sind kostenlos, andere Optionen werden bezuschusst. Zu den Weiterbildungsmöglichkeiten des Superintendents und der Trustees kann leider aufgrund der fehlenden Interviews nichts gesagt werden.

Insgesamt scheint Weiterbildung und Professionalisierung aller Akteure einen hohen Stellenwert zu haben.

## 7.3 Unterstützung der Akteure

Unterstützung der Akteure zeigt sich nach Morrissey (2000) durch strukturelle und soziale Unterstützung, der Unterstützung von Kommunikationsprozessen und anhand von Vertrauen unter den Akteuren.

Hinsichtlich der *strukturellen Unterstützung* stimmen die Lehrpersonen durchschnittlich den Aussagen eher zu, dass bei der Stundenplanfestlegung Zeit ( $n = 21$ ,  $M = 2,90$ ,  $SD = 0,77$ ) und Raum für Teamarbeit vorhanden ist ( $n = 21$ ,  $M = 3,18$ ,  $SD = 0,66$ ).

Die *soziale Unterstützung* und die *Unterstützung von Kommunikationsprozessen* wird durch schulübergreifende und schulinterne Teambildung seitens des Distrikts und der Schule umgesetzt: So gibt es für die Schulleitung, für die Person vom ELL und den/die Special Needs KoordinatorIn Peer-Netzwerke. So treffen sich z.B. die Schulleitungen von Elementary und Junior High Schools aus unterschiedlichen Regionen monatlich im „Network 21“ (SL). Ein weiteres Beispiel sind die vier bis fünf Mal im Jahr stattfindenden Treffen der Personen des ELL zum Erfahrungsaustausch (Interview ELL Teacher).

Innerhalb der Schule gibt es folgende Team (Interview SL, stellvertretende SL, Special Needs KoordinatorIn, Lehrpersonen in spezieller Funktion): das Administration-Team, die Teams der Special Needs KoordinatorIn, der Stufen- und der FachkoordinatorInnen sowie das Instructional Leadership Team (ILT). Bei dem Administration-Team fließen alle Informationen zusammen: denn es besteht aus der Schulleitung, ihrer Stellvertretung, einer Administrationsassistentin sowie den „Leitungen“ aller weiteren Teams. Der/die Special Needs KoordinatorIn koordiniert durch Teamarbeit die Arbeit des/der SozialarbeiterIn, des/der SchulpsychologIn, der vier Personen, die für die verschiedenen Programme zur Förderung der SchülerInnen zuständig sind sowie der Teaching Assistants.

Die Mitglieder des „Instructional Leadership Team“ (ILT), zu welchem die FachkoordinatorInnen, der Special Needs Coordinator<sup>4</sup>, die stellvertretende Schulleitung, die Schulleitung und zwei weitere Lehrpersonen gehören<sup>5</sup>, dienen als AnsprechpartnerInnen für die Lehrpersonen. Sie kontrollieren auch die Umsetzung von Vorgaben (SL). So wird von allen Mitgliedern des ILT erwartet, dass sie zwei bis drei Mal wöchentlich eine Unterrichtseinheit von Lehrpersonen besuchen (SL). Dieses kollegiale Coaching sei nicht an jeder Schule Usus, so die SL. Auch die Lehrpersonen seien dazu angehalten, den Unterricht von KollegInnen mindestens drei bis vier Mal im Jahr zu besuchen, wie auch die Fachleitung bestätigt. Die Schulleitung arbeite zudem mit dem „management by walking around“ (SL). Ungefähr ein Drittel des Tages besuche sie den Unterricht für jeweils zehn bis 15 Minuten, um einen Eindruck zu

---

<sup>4</sup> Der Special Needs Coordinator unterstützt die Arbeit der PsychologIn, der SozialarbeiterIn sowie der Lehrpersonen aus den speziellen Förderprogrammen. Er/sie teilt zudem die SchülerInnen den Förderprogrammen zu.

<sup>5</sup> Letztere gehören nur ausnahmsweise dem ILT an, da sie administrative Aufgaben übernehmen möchten.

gewinnen und AnsprechpartnerIn für Lehrpersonen und SchülerInnen zu sein (SL). In der Schule stehen die Türen zu den Klassenzimmern während des Unterrichts immer offen (eigene Beobachtung).

Als ein Aspekt von Unterstützung gilt zudem das *Vertrauen* (Hord, 1997; Morrissey 2000). Hier soll beispielhaft das Verhältnis zwischen Schulleitung und Assistant Superintendent dargelegt werden, da dieses aufgrund der Kontrollfunktion des Assistant Superintendents gegenüber der Schulleitung besonders interessant ist. Die Schulleitung nennt im Interview zum Netzwerk den/ die Assistant Superintendent als Ansprechperson für Unterstützung. Auch der/die Assistant Superintendent spricht von Unterstützung und nicht von Kontrolle, wenn er über die Beziehung zu den Schulleitungen spricht: „If in fact the school is appearing not to get the results that they ought to be getting with students, then we will increase the amount of support we give that school, not financial support, but personnel support and training“.

An der untersuchten Schule scheint die strukturelle wie soziale Unterstützung gegeben zu sein, Kommunikationsprozesse werden auf horizontaler Ebene durch Teams und Peernetzwerke gefördert und auch die hierarchischen Beziehungen sind im untersuchten Fall von Vertrauen geprägt.

## 8 Diskussion

*Frage 1: Welche Elemente des NPM und der PLG werden in einem „best-case“ de facto umgesetzt?*

Es bestätigt sich die Erwartung, dass in Edmonton Wettbewerbselemente hinsichtlich Ressourcen und Personalfragen weniger durchgesetzt und stattdessen mehr auf Unterstützung und Kooperation gesetzt wird (vgl. DIPF, 2007; van Ackeren, 2010). So existieren hinsichtlich des NPM in Alberta zwar Rankings zwischen den Schulen, die freie Schulwahl spielt wegen der wenigen freien Plätze und der Absprache der Schulen untereinander nur eine geringe Rolle. Eine Selektion der Schulen und eine Effizienzsteigerung durch den Einfluss der KonsumentInnen findet somit in Edmonton nicht statt.

Elemente der Zentralisierung wie die Definition und Überprüfung von Bildungszielen seitens der Regierung erfolgt dagegen, auch wird die Arbeit der Akteure regelmäßig evaluiert. Sanktionen gegen Akteure werden jedoch nicht vorgenommen, stattdessen wird Unterstützung in Form von Weiterbildung angeboten. Zwei Aspekte

sind hier interessant: Erstens haben die Lehrpersonen im Vergleich zu allen anderen Akteuren des Systems keine befristeten Verträge. Zweitens dürfen die Evaluationen aller Beteiligten nicht aufgrund fehlender Sanktionen aus dem Blick geraten. Denn hier könnte die Argumentationsfigur „glaubwürdige Drohung“ (vgl. Osborne 2004) greifen. Glaubwürdige Drohungen bewirken auch ohne Verwirklichung der Drohung antizipatives Verhalten. Hier würde demnach das Wettbewerbsprinzip nicht umgesetzt, es könnte aber indirekt Verhalten beeinflussen – positiv wie negativ.

Hinsichtlich der dezentralen Elemente des NPM obliegen der Schulleitung nur Personalkompetenzen hinsichtlich der Anstellung von BeraterInnen an der Schule. Einstellungen und Kündigungen des Lehrpersonals und auch die Akquisition der BeraterInnen werden hauptsächlich über den Distrikt geregelt.

In Bezug auf die Partizipation an Entscheidungen liegt der Fokus auf den Lehrpersonen und nicht auf den KonsumentInnen wie SchülerInnen und im weitesten Sinne den Eltern. Sie werden gar nicht oder nur sporadisch an Entscheidungen beteiligt. Die Lehrmittel- und Unterrichtsgestaltung obliegt der Provinz oder dem Distrikt. Und die Sachmittelbewirtschaftung wird lediglich zu 10% seitens der Schulleitung bestimmt.

Hinsichtlich der Umsetzung der PLG sind die Charakteristika Kollegialität, Professionalisierung und besonders die Unterstützung der horizontalen Kommunikationsprozesse realisiert. Ein Aspekt wird aber auch in Bezug auf die PLG nicht umgesetzt: Die vertikalen Einflussmöglichkeiten der Akteure sind durch die starken zentralen Elemente der Steuerung (Bildungszielvorgaben etc.) stark eingeschränkt. Von fortlaufenden, systemverändernden Lernprozessen im Sinne der Lernenden Organisation (vgl. Bonsen & Rolff, 2006; Morrissey, 2000) kann nicht gesprochen werden. Die Qualität der Arbeit innerhalb der PLG konnte darüber hinaus in dieser Studie nicht analysiert werden.

### *Frage 2: Welche Vorteile und Probleme ergeben sich aus der Kombination der beiden Konzepte?*

Aus der Kombination der beiden Konzepte wurde sich einerseits eine Stärkung der partizipativen Elemente versprochen. Andererseits wurden drei Spannungsverhältnisse zwischen den Konzepten NPM und PLG beleuchtet: Zentralisierung vs. Partizipation, Output- vs. Inputsteuerung, Wettbewerb vs. Kommunikation.

Der erst genannte Widerspruch zwischen Zentralisierung und Partizipation wird nicht vollkommen aufgelöst: Die Ergebnisse erscheinen auf der strukturellen Dimension wie ein Vexierbild: Einerseits existieren auf unterschiedlichen horizontalen Ebenen, aber auch hierarchieübergreifend Zirkel, in denen auf Partizipation basierend verhandelt wird. Nimmt man die inhaltliche Dimension hinzu, löst sich das Spannungsverhältnis zugunsten von zentralen, hierarchischen Steuerungselementen auf: Die Lehrinhalte und -praxis sind von der zentralen Ebene vorgeschrieben und die Umsetzung wird auch durch vielfältige Methoden auf der Schulebene kontrolliert. Die Autonomie der Mitglieder der PLG wird begrenzt zugunsten der NPM-eigenen Trennung in operative und strategische Ebene (vgl. auch Servage, 2008; Tarnoczi, 2006). Damit lässt sich sagen, dass die zentralen Elemente der Steuerung trotz der partizipatorischen Aspekte überwiegen.

Das zweite Spannungsverhältnis ist hinsichtlich einer Input- und Prozesssteuerung entschieden. Das Output wird gemessen, um diese Steuerungsformen zu optimieren (Gesetze, Abordnungen anzupassen), jedoch nicht, um durch Anreize aufgrund des Outputs Verhalten zu beeinflussen. Die Arbeit der Lehrpersonen wird ebenso wie die Leistungen der SchülerInnen kontinuierlich beobachtet bzw. gemessen. Hier setzen Provinz, Distrikt und Schulleitung bei den SchülerInnen auf groß angelegte Vergleichsarbeiten, kontinuierliche Leistungsüberprüfung – alles während der Schulzeit – und psychologische Tests; bei den Akteuren des Systems wird teilweise jährlich die Arbeit evaluiert, die Arbeit der Lehrpersonen wird durch Monitoring nahezu täglich begleitet. Interessant hierbei ist, dass z.B. das Monitoring der Lehrpersonen (Schulleitung, Distrikt, LehrerInnengewerkschaft) oder SchülerInnen (Provinz, Distrikt, Lehrpersonen, PsychologInnen) unterschiedliche Akteure übernehmen. Die Macht ist auf mehrere Zentren verteilt, die sich gegenseitig kontrollieren („checks and balances“). Die kontinuierliche Weiterbildung aller Akteure (Lehrpersonen, Schulleitung, Assistant Superintendent, Special Needs Coordinator, Person des ELL, ...) stärkt dabei das Professionsverständnis der Akteure und kann als Empowerment der Akteure bezeichnet werden. Dieses Empowerment von Akteuren, die zuvor vielleicht nicht als eigene Professionsgruppe angesehen wurden (z.B. ELL Teacher), verringert Machtasymmetrien in der Zusammenarbeit (z.B. vorher zwischen Verwaltung und Lehrpersonen) und erhöht die Anzahl der Kontrollinstanzen.

Dies birgt ein neues Spannungsverhältnis, denn die voranschreitende Professionalisierung der Akteure kann zur Infragestellung der starken Position der zentralen

Ebene und zu Konflikten zwischen den Professionellen führen. Dieses Konfliktpotential wird durch die hierarchischen Elemente zugunsten der zentralen Ebene aufgelöst, aber vermutlich auch durch die zusätzliche demokratische Legitimation der School Boards abgefedert. Das Expertentum wird sozusagen durch die doppelte demokratische Absicherung in seine Grenzen verwiesen. Dies könnte ein wichtiger Baustein des Bildungssystems sein.

Bei der Frage, überwiegt Wettbewerb oder vertrauensvolle Kommunikation zwischen den Akteuren, zeigen sich eindeutige Ergebnisse: Peer Netzwerke, Teams und z.B. die Rolle des Assistant Superintendents sprechen für eine vertrauensvolle Kommunikationskultur. Wettbewerbselemente wie Konkurrenz unter den Schulleitungen durch die freie Schulwahl werden sogar seitens des Distrikts durch die Initiierung von Netzwerken unter den Schulleitungen („Network 21“) ausgehebelt.

## 9 Fazit

Als Hypothese für weitere Forschung kann aus der Analyse dieses „best-case“ der Bildungsperformanz die folgende Merkmalskombination als möglicherweise erfolgsversprechend abgeleitet werden:

Provinzebene

- Vorgabe von Bildungszielen

Lokale Verwaltungsebene

- Zusätzliche demokratische Legitimation
- Strukturelle und inhaltliche Unterstützung, besonders der Kommunikationsprozesse (PLG) in den Schulen und zwischen Verwaltung und Schule

Schulebene

- Teamarbeit (PLG)

Ebenenübergreifende Merkmale

- Professionalisierung aller Akteure
- System aus „checks and balances“
- kontinuierliches Monitoring der Arbeit aller Akteure
- Kopplungen der Ebenen durch PLG
- Evaluationen ohne Sanktionen, feste Verträge zumindest für das Lehrpersonal

Allerdings lässt eine einzelne Fallanalyse nicht zu, diese Merkmalskombination als „effektiv“ im Sinne einer anzustrebenden Bildungsperformanz zu bezeichnen.

## Literaturverzeichnis

- Achinstein, B. (2002). Conflict amid community: The micropolitics of teacher collaboration. *Teachers College Record*, 104 (3), 421–455. Zugriff am 11.02.2013.
- Alberta Education. (2006). *Professional learning communities: an exploration*, Alberta, Canada. Verfügbar unter <http://www.msdc-mn.org/files-documents/ProfLearnExplore.pdf>.
- Altrichter, H. & Helm, C. (Hrsg.). (2011). Akteure & Instrumente der Schulentwicklung (Professionswissen für Lehrerinnen und Lehrer, Bd. 7). Zürich: Pestalozzianum.
- Altrichter, H. & Rürup, M. (2010). Schulautonomie und die Folgen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch neue Steuerung im Schulsystem* (Educational governance, Bd. 7, S. 111–144). Wiesbaden: VS Verl. für Sozialwissenschaften.
- Bellmann, J. & Weiß, M. (2009). Risiken und Nebenwirkungen Neuer Steuerung im Schulsystem. Theoretische Konzeptualisierung und Erklärungsmodelle. *Zeitschrift für Pädagogik*, 55 (2), 286–308.
- Böhlmark, A. & Lindahl, M. (2007). The impact of school choice on pupil achievement, segregation and costs: Swedish Evidence (Forschungsinstitut zur Zukunft der Arbeit/ Institute for the Study of Labor (IZA), Hrsg.) (Discussion Paper Series Nr. 2786), Bonn
- Bonsen, M. & Rolff, H.-G. (2006). Professionelle Lerngemeinschaften von Lehrerinnen und Lehrern. *Zeitschrift für Pädagogik*, 52 (2), 167–184.
- Burger, J. M., Nadirova, A., Brandon, J., Garneau, B. & Gonnet, C. (2011). Internal and external accountability: Building evidence-informed leadership capacity at all system levels. In T. Townsend & J. MacBeath (Hrsg.), *International handbook of leadership for learning* (S. 243–266). Springer.
- Deutsches Institut für Internationale Pädagogische Forschung (DIPF) (Hrsg.). (2007). Schulleistungen und Steuerung des Schulsystems im Bundesstaat: Kanada und Deutschland im Vergleich. Münster u.a.: Waxmann.
- Dooner, A.-M., Mandzuk, D. & Clifton, R. A. (2008). Stages of collaboration and the realities of professional learning communities. *Teaching and teacher education*, 24, 564–574.
- Eder, F. (1998). Linzer Fragebogen zum Schul- und Klassenklima für die 8.-13. Klasse (LFSK 8-13). Göttingen: Hogrefe.



- Evans, J. E. (1997). *Organizational and individual responses to educational reforms in Alberta* (Alberta University, Hrsg.) (Reports), Edmonton
- Flick, U. (2008). *Triangulation. Eine Einführung* (Qualitative Sozialforschung, Bd. 12, 2. Aufl.). Wiesbaden: Verlag für Sozialwissenschaften.
- Fussangel, K. (2008). *Subjektive Theorien von Lehrkräften zur Kooperation. Eine Analyse der Zusammenarbeit von Lehrerinnen und Lehrern in Lerngemeinschaften*. Verfügbar unter <http://elpub.bib.uni-wuppertal.de/servlets/DerivateServlet/Derivate-1129/dg0802.pdf>.
- Hord, S. M. (1997). *Professional learning communities. Communities of continuous inquiry and improvement*. Austin, Texas: Southwest Educational Development Laboratory (SEDL).
- Jansen, D. (2006). Einführung in die Netzwerkanalyse : Grundlagen, Methoden, Forschungsbeispiele. Opladen: Leske + Budrich
- Leclerc, M., Moreau, A., Dumouchel C. & Sallafranque-St-Louis, F.. Factors that promote progression in schools functioning as professional learning community. *International Journal of Education Policy and Leadership*, 7 (7), 1–14.
- Lomos, C., Hofman, R. H. & Bosker, R. J. (2011). Professional communities and student achievement ... a meta-analysis. *School Effectiveness and School Improvement*, 22 (2), 121–148. Zugriff am 08.02.2013. Verfügbar unter <http://www.tandfonline.com/doi/pdf/10.1080/09243453.2010.550467>.
- Maag Merki, K.; Halbheer, U.; Kunz, A. (2005). Pädagogische Entwicklungsbilanzen an Zürcher Mittelschulen (PEB-ZH). Indikatoren zu Kontextmerkmalen gymnasialer Bildung. Perspektive der Lehrpersonen: Schul- und Unterrichtserfahrungen. Skalen- und Itemdokumentation. FS&S, Pädagogisches Institut, Universität Zürich.
- Marzano, R. J. (2007). Leadership and school reform factors. In T. Townsend (Hrsg.), *International handbook of school effectiveness and improvement* (S. 597–614). Springer.
- Morrissey, M. S. (2000). *Professional Learning Communities. An ongoing Exploration*. Austin, Texas: Southwest Educational Development Laboratory (SEDL). Verfügbar unter <http://allthingsplc.info/pdf/articles/plc-ongoing.pdf>.
- Muno, W. (2009). Fallstudien und die vergleichende Methode. In S. Pickel, G. Pickel, H.-J. Lauth & D. Jahn (Hrsg.), *Methoden der vergleichenden Politik- und Sozialwissenschaft. Neue Entwicklungen und Anwendungen* (S. 113–131). Wiesbaden: VS Verlag für Sozialwissenschaften.

- OECD (2004). *Bildung auf einen Blick. OECD-Indikatoren 2004*. Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- OECD (2006). PISA 2003. Technical report. Paris: OECD.
- OECD (2007). PISA 2006. Science Competencies for Tomorrow's World. Paris: OECD.
- OECD (2010a). Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften (PISA 2009 Ergebnisse, Bd. 1). Bielefeld: Bertelsmann.
- OECD (2010b). TALIS 2008. Technical report. Paris: OECD.
- Osborne, M. J. (2004). *An introduction to game theory*. Oxford: Oxford University Press.
- Rolff, H.-G. (2001). Professionelle Lerngemeinschaften. Eine wirkungsvolle Synthese von Unterrichts- und Personalentwicklung. In H. Buchen & L. Horster (Hrsg.), *Schul-leitung und Schulentwicklung Erfahrungen, Konzepte, Strategien* (D 6.5). Berlin: Raabe Fachverlag für Bildungsmanagement.
- Sackney, L. (2007). History of the school effectiveness and improvement movement in Canada over the past 25 years. In T. Townsend (Hrsg.), *International handbook of school effectiveness and improvement* (S. 167–182). Springer.
- Sackney, L. (2011). Leadership and learning: the Canadian context. In T. Townsend & J. MacBeath (Hrsg.), *International handbook of leadership for learning* (S. 51–56). Springer.
- Schedler, K. & Proeller, I. (2006). New Public Management (UTB 2132 : Public Management Betriebswirtschaft, 3. Aufl.). Bern: Haupt.
- Schwarzer, R. & Jerusalem, M. (Hrsg.). (1999). Skalen zur Erfassung von Lehrer und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen. Berlin: Freie Universität Berlin und Humboldt-Universität zu Berlin.
- Servage, L. (2008). Critical and transformative practices in professional learning communities. *Teacher Education Quarterly*, 35 (1).
- Tarnoczi, J. (2006). Critical reflections on professional learning communities in Alberta. *Electronic Journal of Sociology*. Verfügbar unter: <http://www.sociology.org/content/2006/tier2/tarnoczi.html>; Stand: 10.12.12
- Thillmann, K. (2011). Reorganisation von Schulen unter den Bedingungen der Neuen Steuerung. In F. Dietrich, M. Heinrich & N. Thieme (Hrsg.), *Neue Steuerung - alte Ungleichheiten? Steuerung und Entwicklung im Bildungssystem* (S. 65–77). Münster: Waxmann.

- Thompson, S. C., Gregg, L. & Niska, J. M. (2004). Professional learning communities, leadership, and student learning. *Research in Middle Level Education Online*, 28 (1), 1–15.
- Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *The Journal of Political Economy*, Vol. 64, No. 5. (Oct., 1956), S. 416-424.
- Tomlinsen, S. (2000). Wie wirken sich Bildungsmärkte auf ethnische Minderheiten aus? In F.-O. Radtke & M. Weiß (Hrsg.), *Schulautonomie, Wohlfahrtsstaat und Chancengleichheit. Ein Studienbuch* (S. 201–224). Opladen: Leske + Budrich.
- Van Ackeren, I. & Brauckmann, S. (2010). Internationale Diskussions-, Forschungs- und Theorieansätze zur Governance im Schulwesen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch neue Steuerung im Schulsystem* (Educational governance, Bd. 7, S. 41–61). Wiesbaden: VS Verl. für Sozialwissenschaften.
- Von Hayek, F. A.. (1969). Der Wettbewerb als Entdeckungsverfahren. Walter Eucken Institut Freiburg i. Br.. Freiburger Studien – Gesammelte Aufsätze von F.A. von Hayek. Tübingen: J.C. B. Mohr.
- Wößmann, L. (2008): Zentrale Abschlussprüfungen und Schülerleistungen. Individualanalysen anhand von vier internationalen Tests. *Zeitschrift für Pädagogik*, 54 (6), S. 810-826.

## Internetquellen

Alberta Education (2012):

<http://www.education.alberta.ca/admin/testing/achievement.aspx>, Zugriff: 29.05.2012

EPSB (2009a): Trustee Electoral Ward Boundaries. Verfügbar unter:

[http://www.epsb.ca/board/dec08\\_09/item06.pdf](http://www.epsb.ca/board/dec08_09/item06.pdf), Zugriff: 29.05.2012

EPSB (2009b): Edmonton Public Schools Board Policies and Regulations. Topic: Supervision and Evaluation of Teachers:

<http://local37.teachers.ab.ca/SiteCollectionDocuments/Local37/2009-2010/Notices%20and%20registration%20forms/Rpt%20-%20Joint%20PD%20Framework%20Attachment%20V.pdf>, Zugriff: 29.05.2012

EPSB (2010): <http://www.epsb.ca/about/districtoverview.shtml>; Zugriff: 31.02.2010

Province of Alberta (2011): School Act. Revised Statutes of Alberta 2000:

[http://www.qp.alberta.ca/574.cfm?page=s03.cfm&leg\\_type=Acts&isbncln=9780779733941](http://www.qp.alberta.ca/574.cfm?page=s03.cfm&leg_type=Acts&isbncln=9780779733941), Zugriff: 29.05.2012

Schmold, S. (2010): Improving High School Completion. Alberta School Board Perspectives. Alberta School Board Association (ASBA):

[http://www.asba.ab.ca/files/pdf/high\\_school\\_completion\\_report.pdf](http://www.asba.ab.ca/files/pdf/high_school_completion_report.pdf), Zugriff: 08.02.13

Anhang (Erklärung, Lebenslauf)

### **Erklärung**

Hiermit erkläre ich,  
dass die Dissertation von mir selbst ohne unerlaubte Beihilfe verfasst worden ist und  
diese Dissertation noch an keiner anderen Fakultät eingereicht wurde.

Dortmund, 25.6.2014

Ort, Datum und Unterschrift

D. Jg

## **Curriculum Vitae – Daniela J. Jäger**

geb. 19.02.1979, Viersen, Deutschland, 1 Sohn

### **Forschungsinteressen**

- Educational Governance
- Schulqualität und -entwicklung
- Ganztagschulen
- Zentrale Abschlussprüfungen
- Quantitative Netzwerkanalyse

### **Lebenslauf**

- Technische Universität Dortmund, Institut für Schulentwicklungsforschung, seit 2013
- Verteidigung der kumulativen Dissertation, 2013
- Forschungsaufenthalt in Kanada, Fallstudien in vier Provinzen, Januar-Mai 2010
- Universität Zürich, Institut für Erziehungswissenschaft, 2009-2013
- Pädagogische Hochschule Freiburg, Institut für Erziehungswissenschaft, 2008-2009
- Universität zu Köln, Lehrstuhl Vergleichende Politikwissenschaft, 2007-2009
- Universität zu Köln: Studium der Politologie, Philosophie und Deutschen Philologie Abschluss: Magistra Artium, 2000-2006
- Rheinisch-Westfälische Technische Hochschule Aachen: Studium der Politologie, Philosophie und Arbeits- und Betriebspsychologie, 1998-2000
- Abitur am Albertus-Magnus-Gymnasium, Viersen-Dülken, 1998